Volume 31 Number 1 March 1980
BJPIA5 30 (No. 1) 1-109 (1980) ISSN 0007-0882

The British Journal for the Philosophy of Science

22

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: JOHN WORRALL

Subscriptions

The *Journal* is published quarterly. The subscription price for mailing addresses in Britain is £7.00 net (U.S.A. and Canada U.S. $22.00, other overseas £11.00) per volume post free; single issues £2.50 net (U.S.A. and Canada U.S. $8.00, other overseas £4.00) plus postage. Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

For members of the British Society for the Philosophy of Science—who should order direct from Aberdeen University Press—the subscription price is £4.50 (U.S.A. and Canada U.S. $14.00, other overseas £7.00). Membership of the Society is international and is by no means confined to people resident in the United Kingdom.

Vol

Microform

The current volume of the *J* ble in microform to subscribers to the printed version.

Back Volumes

Volumes 1–16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17–30 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application. Volumes 1–29 are also available in microform from Kraus Reprint Corporation.

Typescripts and Review Books

Typescripts, correspondence and books for review should be sent to the Editor at the Department of Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street, London WC2A 2AE.

Review books are accepted on the understanding that, although they will appear in the 'Recent Publications' list, a full review cannot be guaranteed, and that books not reviewed are not returnable.

Notes to Contributors

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Detailed instructions on the lay-out of typescripts are published occasionally at the end of the *Journal* and anyway may be obtained from the Editor.

Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

© British Society for the Philosophy of Science 1980

For permission to reproduce material from *The British Journal for the Philosophy of Science,* please apply to the Editor.

CUK-H00681-22-P3333

The British Journal for the Philosophy of Science

Volume 31

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press
1980

P2223

CONTENTS

ARTICLES

page

1 EINSTEIN, MEYERSON AND THE ROLE OF MATHEMATICS IN PHYSICAL DISCOVERY
Elie Zahar
45 BAYESIANISM VERSUS BACONIANISM IN THE EVALUATION OF MEDICAL DIAGNOSES
L. Jonathan Cohen
111 BRAIN STATES
Kathleen Wilkes
131 BAYESIAN CONDITIONALISATION AND THE PRINCIPLE OF MINIMUM INFORMATION
P. M. Williams
145 MODELS IN PHYSICS
Michael Redhead
165 ON THE INTERPRETATION AND TRANSITIVITY OF NON-STANDARD SYNCHRONISMS
R. Francis
213 A NEW *Begriffsschrift* (I)
John P. Mayberry
255 CONVENTIONALISM ABOUT SPACE AND TIME
Richard Swinburne
317 SOCIAL PROPENSITIES
Peter Urbach
329 A NEW *Begriffsschrift* (II)
John P. Mayberry

DISCUSSIONS

63 POPPERIAN LANGUAGE-ACQUISITION UNDEFEATED
Geoffrey Sampson
175 SOME COMMENTS ON THIRD WORLD EPISTEMOLOGY
L. Jonathan Cohen
273 ZAHAR ON MACH, EINSTEIN AND MODERN SCIENCE
Paul Feyerabend
282 ON THINGS AND CAUSES IN SPACETIME
Hugh Mellor

REVIEW ARTICLES

69 ARMSTRONG'S THEORY OF UNIVERSALS (review of D. M. Armstrong: *Universals and Scientific Realism*)
David H. Sanford
181 THE HALT AND THE BLIND: PHILOSOPHY AND HISTORY OF SCIENCE (review of C. Howson (*ed.*): *Method and Appraisal in the Physical Sciences*)
Thomas Kuhn

page

289 PHENOMENOLOGY AND THE INFINITE IN MATHEMATICS (review of F. Kaufmann: *The Infinite in Mathematics*)
Donald Gillies

359 OBJECTIVITY, RATIONALITY, INCOMMENSURABILITY AND MORE (review of T. S. Kuhn: *The Essential Tension*)
Harvey Siegel

375 REICHENBACH'S THEORY OF REASONABLE ASSERTION (review of H. Reichenbach: *Laws, Modalities and Counterfactuals*)
Evan Jobe

REVIEWS

CARROLL, LEWIS: *Symbolic Logic* review by Irving Copi, 81

MANIER, E.: *The Young Darwin and His Cultural Circle* review by M. J. S. Hodge, 85

PUTNAM, HILARY: *Meaning and the Moral Sciences* review by Robert Nola, 91

HUTCHISON, T. W.: *Knowledge and Ignorance in Economics* review by David Papineau, 98

MATTESSICH, RICHARD: *Instrumental Reasoning and Systems Methodology* review by D. C. Phillips, 89

WILLIAMS, B.: *Descartes: The Project of Pure Enquiry* and E. Curley: *Descartes Against the Skeptics* review by Ronald Rubin, 104

HOLTON, GERALD: *The Scientific Imagination: Case Studies* review by Noretta Koertge, 193

BUTTS, R. E. and PITT, J. C. (*eds*). *New Perspectives on Galileo* review by C. B. Schmitt, 195

ULLMAN-MARGALIT, E.: *The Emergence of Norms* review by Nigel Howard, 199

DUMMETT, MICHAEL: *Elements of Intuitionism* review by Peter Eggenberger, 299

SAARINEN, ESA (*ed.*): *Game-theoretical Semantics* review by Pavel Tichy, 301

WILSON, M. D.: *Descartes* review by Harry Bracken, 307

EARMAN, J. S., GLYMOUR, C. N. and STACHEL, J. J. (*eds.*): *Foundations of Space-Time Theories* review by Roger Jones, 311

FEYERABEND, PAUL: *Science in a Free Society* review by Noretta Koertge, 385

GREGORY, FREDERICK: *Scientific Materialism in Nineteenth Century Germany* review by Hans Sluga, 390

HOOKWAY, C. and PETTIT, P. (*eds*): *Action and Interpretation* review by Ian Jarvie, 396

SALMON, WESLEY (*ed.*): *Hans Reichenbach: Logical Empiricist* review by Richard Swinburne, 401

HUFF, D. and PREWETT, O. (*eds*): *The Nature of the Physical Universe* review by Michael Redhead, 404

CRAIG, W. L.: *The Kalam Cosmological Argument* review by Gerald Whitrow, 408

RESCHER, N.: *Cognitive Systematisation* review by Susan Haack, 411

TORRETTI, ROBERTO: *Philosophy of Geometry from Riemann to Poincaré* review by J. L. Lucas, 414

page

MISCELLANEOUS

109 NOTES TO CONTRIBUTORS
204 RECENT PUBLICATIONS
211 FIFTH INTERNATIONAL WITTGENSTEIN SYMPOSIUM
212 BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE: Annual Conference, 1980
316 CONFERENCE ON SPACETIME AND CAUSALITY

INDEX OF AUTHORS

ARTICLES

Cohen, L. J. Bayesianism versus Baconianism in the Evaluation of Medical Diagnoses, 45
Francis, R. On the Interpretation and Transitivity of Non-standard Synchronisms, 165
Mayberry, J. P. A New *Begriffsschrift* (I), 213
Mayberry, J. P. A New *Begriffsschrift* (II), 329
Redhead, M. Models in Physics, 145
Swinburne, R. Conventionalism about Space and Time, 255
Urbach, P. Social Propensities, 317
Wilkes, K. V. Brain States, 111
Williams, P. M. Bayesian Conditionalisation and the Principle of Minimum Information, 131
Zahar, E. G. Einstein, Meyerson and the Role of Mathematics in Physical Discovery, 1

DISCUSSIONS

Cohen, L. J. Some Comments on Third World Epistemology, 175
Feyerabend, P. Zahar on Mach, Einstein and Modern Science, 273
Mellor, H. D. On Things and Causes in Spacetime, 282
Sampson, G. Popperian Language-Acquisition Undefeated, 63

REVIEW ARTICLES

Gillies, D. Phenomenology and the Infinite in Mathematics (review of F. Kaufmann: *The Infinite in Mathematics*), 289
Jobe, E. Reichenbach's Theory of Reasonable Assertion (review of H. Reichenbach: *Laws, Modalities and Counterfactuals*), 375
Kuhn, T. S. The Halt and the Blind: Philosophy and History of Science (review of C. Howson (*ed.*): *Method and Appraisal in the Physical Sciences*), 181
Sanford, D. H. Armstrong's Theory of Universals (review of D. M. Armstrong: *Universals and Scientific Realism*), 69
Siegel, H. Objectivity, Rationality, Incommensurability and More (review of T. S. Kuhn: *The Essential Tension*), 359

REVIEWS

Bracken, H. M. D. Wilson: *Descartes*, 307
Copi, I. Lewis Carroll: *Symbolic Logic*, 81
Eggenberger, P. Michael Dummett: *Elements of Intuitionism*, 299
Haack, S. N. Rescher: *Cognitive Systematisation*, 411
Hodge, M. J. S. E. Manier: *The Young Darwin and His Cultural Circle*, 85
Howard, N. E. Ullman-Margalit: *The Emergence of Norms*, 199
Jarvie, I. C. Hookway and P. Pettit (*eds*): *Action and Interpretation*, 396
Jones, R. J. S. Earman, C. N. Glymour and J. J. Stachel (*eds*): *Foundations of Spacetime Theories*, 311
Koertge, N. Paul Feyerabend: *Science in a Free Society*, 385
Koertge, N. Gerald Holton: *The Scientific Imagination: Case Studies*, 193

LUCAS, J. L.	Roberto Torretti: *Philosophy of Geometry from Riemann to Poincaré*, 414
NOLA, R.	Hilary Putnam: *Meaning and the Moral Sciences*, 91
PAPINEAU, D.	T. W. Hutchison: *Knowledge and Ignorance in Economics*, 98
PHILLIPS, D. C.	Richard Mattessich: *Instrumental Reasoning and Systems Methodology*, 89
REDHEAD, M.	D. Huff and O. Prewett (*eds*): *The Nature of the Physical Universe*, 404
RUBIN, R.	B. Williams: *Descartes: The Project of Pure Enquiry* and E. Curley: *Descartes Against the Skeptics*, 104
SCHMITT, C. B.	R. E. Butts and J. C. Pitt (*eds*): *New Perspectives on Galileo*, 195
SLUGA, H.	Frederick Gregory: *Scientific Materialism in Nineteenth Century Germany*, 390
SWINBURNE, R.	Wesley Salmon (*ed.*): *Hans Reichenbach: Logical Empiricist*, 401
TICHY, P.	Esa Saarinen (*ed.*): *Game-theoretical Semantics*
WHITROW, G.	W. L. Craig: *The Kalam Cosmological Argument*, 408

Brit. J. Phil. Sci. 31 (1980), 1–43 *Printed in Great Britain*

Einstein, Meyerson and the Role of Mathematics in Physical Discovery

by ELIE ZAHAR

1 *The Role of Mathematics within Physics*
  (*i*) *Increase of Content Through Translation into a Mathematical Language*
  (*ii*) *Realistic Interpretation of Mathematical Entities*
2 *Meyerson's Principle of Identity*
  (*i*) *Legal Explanations*
  (*ii*) *Causal Explanations and Conservation Laws*
  (*iii*) *The Geometrisation of Nature*
3 *The Continuity Between* STR *and* GTR
  (*i*) *The Heuristic Role of the Covariance Principle*
  (*ii*) *The Abandonment of Special Relativity*
  (*iii*) *The Heuristic Role of the Principle of Equivalence*
  (*iv*) *The Field Equations*
  (*v*) *The Negative Heuristic*
Appendix *The Status of the Relativity Principle*

INTRODUCTION

This paper addresses itself to three seemingly distinct but in fact interrelated issues: the role of mathematics in physical discovery, the heuristic role of certain philosophical ideas and the problem of the continuity between the Special and the General Theories of Relativity (henceforth referred to as *STR* and *GTR* respectively). The experimental confirmations of *GTR*: precession of Mercury's perihelion, bending of the light rays and red shift, are taken to be well-known. The problem to be tackled is whether this empirical success can be claimed by a single research programme comprising both *STR* and *GTR*. These two theories seem to be very different scientific hypotheses (in fact, according to some physicists, the empirical status of *GTR* is doubtful). Whittaker, for example, distinguishes between the Relativity Principle which he attributes largely to Lorentz and Poincaré, and *GTR* which he regards as a new theory of gravitation. In fact, *prima facie*, there should be no direct connection between the

Received 15 *November* 1978

Relativity Principle on the one hand and any particular physical phenomenon like gravitation on the other. In his [1905] Einstein develops a new kinematics without referring to any specific physical theory; he then goes on to formulate Mechanics and Electrodynamics in a Lorentz-covariant way. The kinematical section of his paper remains logically, if not heuristically, independent of the subsequent sections. One might have expected the same to happen in *GTR*: namely the formulation of some new kinematics, followed by a gravitational theory set against an already articulated geometrical background. Nothing of the sort happens: gravitation enters the discussion right from the start.

This is the first reason for posing the continuity problem; the second one is that the heuristics of the programme experienced a major shift during the period which led to *GTR*. The methods applied by Einstein, Planck, Tolman and Lewis before 1910 differ widely from those used after 1910 by Einstein, Kaluza, Weyl, Eddington and Schrödinger say. Whereas in his [1905] Einstein painstakingly tries to give operational definitions of duration and length, in the [1916] paper the coordinates have lost any direct physical meaning. This is a far cry from the Einstein who, according to Bridgman (see his [1936]), replaced absolute time and space by clock-time and rod-space. Whereas in their [1909] Tolman and Lewis base their new definitions of momentum on a consideration of how different observers *measure* distance and time, Eddington in his [1924] sets out to geometrise physics *a priori*. From being experimental scientists mainly concerned with actual physical measurements, the Relativists had become *a priorist* geometers. The whole tenor of their approach seemed to have undergone a major change. This change was of course reflected at the experimental level. While *STR* is a highly confirmed empirical theory, many scientists like Trautmann wonder whether *GTR* belongs to Physics or to Pure Mathematics. Whatever the case may be, there is an obvious disparity between the complex mathematical machinery of *GTR* and the paucity of its experimental results. This throws into relief, on the one hand, the heuristic role of mathematics within physics and, on the other, the question of whether the development leading to *GTR* occurred within a single programme comprising both the Special and the General Theories.

1 THE ROLE OF MATHEMATICS WITHIN PHYSICS

In the preface of their [1924] Hilbert and Courant express their regret that, towards the beginning of the twentieth century, the fruitful interplay between mathematics and physics had almost stopped. They write:

> Since the Seventeenth Century physical intuition has served as a vital source for mathematical problems and methods. Recent trends and fashions have however weakened the connection between mathematics and physics; mathematicians, turning away from the roots of mathematics in intuition, have concentrated on refinement and emphasised the postulational side of mathematics, and at times have overlooked the unity of their science with physics and other fields. In many cases physicists have ceased to appreciate the attitudes of mathematicians. This rift is unquestionably a serious threat to science as a whole . . .

It is well-known that empirical science has often exerted a beneficial influence on the development of mathematics. Physics sets problems for which mathematical solutions are urgently required; as a result certain branches of pure mathematics receive a powerful impetus and develop rapidly. For a long time mathematics and physics were regarded as identical disciplines; or rather people were not clear about the differences between them. Euclidean geometry for example was taken to describe the properties of real, albeit idealised, space. Its postulates, as distinct from the axioms, were accepted because they supposedly expressed self-evident truths about physical space. In Einstein's words ([1934], p. 119), geometry constituted one of the oldest physical theories. In the preface to his *Principia* Newton treats geometry as a branch of mechanics, *i.e.* as a branch of physics:

> Therefore geometry is founded in mechanical practice and is nothing but that part of universal mechanics which accurately proposes and demonstrates the art of measuring.

In many cases physics played the dominant role in that it was physics which gave rise to the mathematical problems and to a large extent dictated the type of mathematics to be used. In his Theory of Fluxions, Newton constructed the Calculus specifically for the study of differentiable motions: the fluent variable was the time and the fluxion the instantaneous velocity.[1] Hence analysis, *i.e.* the discipline which dominated mathematical thinking for over two centuries, owes its origins to physics. Similarly, the study of differential equations and the development of what is now known as Advanced Calculus were closely connected with the articulation of the Newtonian Programme in the eighteenth century. A similar process took place in the nineteenth century when Faraday, using 'line of force' as a new physical concept, enunciated laws of which Maxwell later gave a mathematical formulation. This powerfully contributed to the development of vector algebra and vector analysis. These examples thus illustrate the heuristic role of physics with regard to mathematics; in each of the cases so far considered it was physics which led to the discovery of an appropriate mathematical theory. It looks as if, whenever mathematics and physics have interacted, the latter set the pace. Has this always been the case? I would like to examine whether the reverse process has ever taken place, *i.e.* whether the solution of a mathematical problem has ever led to the discovery of a physical theory. *Prima facie*, this process ought to be possible. There is no reason why the mathematical rendering of some vague physical principle should not unexpectedly capture more of the truth than the scientist was initially aware of. This question concerns the heuristic role of mathematics *vis-à-vis* physics. There are two important ways in which mathematics further physical discovery.

---

[1] See Oskar Becker [1963], pp. 144–67.

(*i*) *Increase of Content through Translation into a Mathematical Language*

As mentioned above, the scientist may start from an intuitive physical principle. Through being translated into some mathematical theory available at the time, the principle is generally modified: it may in particular acquire some surplus structure[1] and thus become a stronger physical assumption. For example, Fresnel set out to give a mathematical formulation to his conjecture that light is a wave process through the ether. He instinctively resorted to the periodic function with which he was most familiar, namely sin x. His original assumption that light is a wave phenomenon is obviously weaker than his final hypothesis that the wave is representable by the function $\sin(2\pi t/T)$.

It has often happened that certain physical hypotheses which require very specific mathematical theories for their formulation were proposed after, and sometimes immediately after, the mathematical theories were elaborated. This seemingly pre-established harmony between the mathematics and the subsequent physics has been a cause for puzzlement among historians and philosophers of science. One example of such a coincidence is the relationship between the mathematics of Hilbert space and the formulation of Quantum Mechanics. Another very good example is the development of Riemannian Geometry prior to the emergence of *GTR*. It was as if Gauss, Riemann, Ricci and Levi-Civita had Einstein in mind when constructing their non-Euclidean geometries. Einstein himself considered his discoveries as a continuation of Gauss's and Riemann's work. Yet Einstein was a physicist and Riemann a pure mathematician. In his 'Geometry and Experience' Einstein strictly distinguished between geometry *qua* mathematical system on the one hand and geometry *qua* physical theory on the other. It is *prima-facie* strange that mathematicians should anticipate the work of later physicists. The solution of this puzzle is two fold. On the one hand there is no pre-established harmony between mathematics and physics. The physicist forces, or impresses, his principles into an existing mathematical framework. The physical principles thus acquire the surplus structure of the mathematical system which is used. When Einstein decided he could not account for gravity in Special Relativistic terms, he turned to his friend Grossmann, a working mathematician steeped in the Riemannian tradition. It is small wonder that Grossmann used Riemannian geometry to cast the field equations for empty space in the form $R_{ij} = 0$. In this way Grossman and Einstein geometrised gravitation. The harmony between mathematics and physics is thus forcibly established, not *pre*-established. On the other hand, however, Riemannian geometry had to be of such a nature as to lend itself to the use to which it was put by Einstein and Grossmann. In other words, this geometry had to have an empirical or a quasi-empirical character which allowed physics to be imbedded in it.

---

[1] For the role which mathematical surplus structure plays in physics, see M. L. G. Redhead [1975].

That this was in fact the case is clearly indicated by the following quotation from Riemann's famous dissertation 'About the hypotheses which lie at the basis of geometry':

> It will emerge from this that a multiply extended magnitude [*Grösse*] is capable of containing within itself different metric relations, and hence that space constitutes a special case of a three dimensional magnitude [*dreifach ausgedehnte Grösse*]. From this it necessarily follows that geometrical propositions are not deducible from the concept of a magnitude, but that those properties, through which space is distinguished from other conceivable three-dimensional magnitudes, are derivable only from experience. These facts, like all facts, are not necessary, they possess only empirical certainty, they are hypotheses; one can therefore examine their probability which is, within observational limits, indeed very high. ([1953], pp. 272–3; my translation.)

Riemann's approach to geometry is quasi-empirical in that it is based on the notion of measured distance between neighbouring points. Riemann physicalised mathematical geometry, thus making it possible for Einstein and Grossman to geometrise gravitation. We should however remark that empirical considerations alone, *i.e.* without idealising *a priori* assumptions, do not determine the structure of Riemannian Geometry. Riemann himself realised that the differentiability condition is not derivable from experience:

> I will undertake this task only under certain restrictive assumptions and shall first confine myself to those lines on which the ratios between the magnitudes *dx* . . . vary continuously. The question of the validity the geometrical presuppositions within infinitely small regions hangs together with the question about the foundations of all metrical relationships in space. To this question, which can be reckoned to the science of space, the above remark applies, namely

that, for a discrete manifold, the principles of the metric relationships are already contained in the concept of the manifold itself, whereas, in the case of a continuous manifold, these principles have to be imported from outside (*ibid.*, p. 286).

There is another assumption, of whose *a priori* character Riemann seems to be unaware, namely that space is determined by a metric, *i.e.* by one fundamental symmetric tensor g_{mn}. That this is an *a priori* assumption is shown by the fact that, after 1916, purely affine geometries were developed which could accommodate the facts then known to Einstein.

Let me say that I am by no means committed to the Kantian notion of the *a priori*. I use '*a priori*' in the sense of 'not imposed by experience'. This conception of the *a priori* is perfectly compatible with Mach's. Mach[1] identified *a priori* knowledge with the instinctive, innate knowledge which may have arisen through the gradual adaptation of the mind to its surroundings, *i.e.* through natural selection. Thus, for Mach, *a priori* knowledge is inheritable, *i.e.* genetically imbedded in the organism.

Going back to Riemann, we see that metricity and differentiability are *a priori* assumptions which belong to the surplus structure of geometry with respect to physics. Thus, forcing physics into the Riemannian mould imposed more constraints than are warranted by observational results. Does this imply that the later hypotheses of Weyl, Eddington and Schrödinger were more 'empirical', or more 'inductive', than *GTR*? By no means. Metricity was not replaced by laws induced from experience, but by different *geometrical* assumptions. Weyl[2] links the presence of the electromagnetic field to the change in *length* of a vector transported parallel to itself. Eddington proposes to explain gravitation and electromagnetism by means of the symmetric and anti-symmetric parts of the Ricci *tensor*.[3] As is well known, the Ricci tensor is obtained by *contraction* from the curvature tensor B^m_{ijn}; and B^m_{ijn} is obtained from the affinity Γ^m_{ij} by considering the variation of a vector taken round an infinitesimal loop. Eddington derives the symmetry in u and v of Γ^m_{uv} from a *commutativity* condition. Schrödinger[4] constructs his theory by applying a *variational principle* to a certain integral; where the integrand is what he and Eddington consider as the simplest *invariant density* built out of the curvature tensor, namely $\sqrt{-det[R_{ij}]}$. I mention all these details in order to make the following point: length, parallel transport, contraction, commutativity, tensor densities and variations of integrals are *mathematical* notions not directly abstracted from experience. The physicist operates at the mathematical level, hoping that his operations mirror certain features of reality. However, he is not very clear as to how this mirroring takes place, so he lets himself be guided by the syntax, or by the symbolism, of some mathematical system.

---

[1] See Mach [1883], Introduction. Also [1905], pp. 164–82.

[2] See Weyl [1923], § 40.

[3] See Eddington [1924], § 97.

[4] See Schrödinger [1950], chapter 12.

Of course, these seemingly abstract unified field theories have novel empirical consequences. It is precisely because Weyl's hypothesis had some undesirable physical consequences that Einstein rejected it. This illustrates the point already made that physics is strengthened by mathematical surplus structure.

Mathematical speculation thus seems to have played a dominant role in the development of Einstein's programme. However, it can also be said that the problems which physics faced after 1916, *e.g.* the fact that *GTR* does not treat gravitation and electromagnetism on a par, gave rise to new types of non-Riemannian geometry. Can we not say with equal right that physics played a dominant heuristic role with regard to mathematics? In my opinion the relationship between mathematics and physics is best described in dialectical terms as a to and fro movement between two poles. One moves from physical principles to idealising mathematical assumptions; then back to some more physics; then forward to fresh mathematical innovations with ever increasing surplus structure. The so called harmony between physics and mathematics is not a miracle but the result of an arduous process of mutual adjustment.

(*ii*) *Realistic Interpretation of Mathematical Entities*

There is a second way in which mathematics can play a fundamental role in physical discovery. The usual method in theoretical physics is to give mathematical expression to some physical hypothesis, then use logico-mathematical techniques to draw consequences from the hypothesis. In doing this, the physicist may resort to a number of mathematical operations; these are sometimes in the nature of 'tricks' or 'gimmicks' which simplify the deductive process. Duhem[1] pointed out that it would be foolish to insist on giving a physical interpretation to *all* mathematical quantities and operations used in a scientific theory. Duhem is evidently right: adding lengths corresponds to juxtaposing rods; multiplying lengths may correspond to the construction of rectangular areas but multiplying t by $\sqrt{-1}$, though convenient, does not appear to be susceptible of any physical interpretation. However, through trying to find a *realistic* interpretation of certain mathematical entities which seem at first sight to be devoid of any physical meaning, the scientist may be led to a physical conjecture.[2] This can occur in two different ways.

---

[1] See Duhem [1906], part 2, chapter 1.

[2] For the important role which philosophical realism plays in the logic of discovery, see Popper [1963], pp. 97–119.

The first and straightforward way is through an increase in the empirical content of a given theory. By being realistically interpreted, part of the mathematical scaffolding becomes physically meaningful, hence testable, at least in principle. The theory remains syntactically unchanged, but its observational content, *i.e.* its contact with 'observable' reality, is extended. Dirac's equation for the electron, his explanation of spin and especially his discovery of the positron are cases in point. Dirac proposed a relativistic equation which was found to possess negative energy solutions. *Prima facie* such solutions cannot be physically interpreted. By insisting on interpreting the negative solutions, Dirac predicted the existence of the positron; the absence of an electron of charge $-e$ and energy $-E$ was interpreted as the presence of a positron, that is of a particle of charge $+e$ and energy $+E$.

The second way in which the realistic interpretation of mathematical entities leads to new discoveries is both more complex and more interesting than the previous one. The theory itself is altered in this process; that is, the syntactical expression of the physical principles themselves does not escape unscathed. This usually occurs as follows. We start from some hypothesis H which we may want to modify for reasons we need not go into. We express H in an equivalent form,[1] $H^*(t)$, which brings out a certain mathematical entity t. t is then given a realistic interpretation which subsumes it under a philosophical category, *e.g.* the category of substance; this category obeys some very general laws, *e.g.* conservation or symmetry laws. It is then found that $H^*(t)$ violates these laws. The breakthrough is achieved when $H^*(t)$ is modified into $H'(t)$ which conforms to the laws in question. We shall see that, by using variational methods, Einstein re-expressed his original field equations F in the form $F^*(t)$, where t was interpreted as a gravitation energy matrix, despite the fact that t is frame-dependent. Let T be the non-gravitational energy tensor. Thus $(T+t)$ must obey a conservation law and the field equations must be symmetric in T and t. Einstein noticed that $F^*(t)$ violates this condition; he modified $F^*(t)$ so as to obtain equations symmetric in T and t.

2 MEYERSON'S PRINCIPLE OF IDENTITY

In this section I shall give a brief exposition of Meyerson's philosophy of science. I shall argue that this philosophy accounts *in terms of a single principle* both for the applicability of mathematics to physics and for the role which conservation laws and certain philosophical notions like that of substance play in the progress of science. This single Meyersonian principle allows of many formulations which, though distinct, bear to one another a sort of family resemblance. According to Meyerson,[1] all explanations, whether scientific or commonsensical, spring from one basic tendency of the human mind; namely the tendency to deny diversity and change; or to assert the existence of constants behind the fleeting appearances; or to explain the Many in terms of the One; or to subsume the flux of Becoming under the immutability of Being. The best formulation, in my opinion, is that the human mind inevitably tends to deny diversity and assert sameness or identity both in space and in time. For the human mind, only the undifferentiated One is real, everything else is appearance. This sounds like the worst kind of metaphysical doctrine which is bound to be of no relevance to science. May I however ask the reader to be patient, suspend disbelief for the time being and let me examine how much of the role which mathematics and conservation laws play in physics can be derived from Meyerson's principle of identity? Note that I have spoken of this principle as of a tendency inherent in the human mind. If this tendency went unchecked, it would assert the existence of a unique Parmenidean Sphere which is undifferentiated in space and changeless in time; it would thus degrade all phenomena to the rank of mere sensory illusions. Becoming would thus be mere illusion and only Being would be real. The whole of science can be conceived as a sequence of attempts to rescue as much of Parmenides's sphere as is compatible with the phenomena. It is well-known that Greek Atomism postulates the void together with a plurality of particles each of which can be assimilated to a Parmenidean sphere. Both the atoms and the void are immutable in themselves. The flux of appearances is explained in terms of different configurations of the same particles within the same empty space. Let A and K stand for two consecutive states of affairs which, following Meyerson, will henceforth be referred to as the antecedent and the consequent respectively. Thus, according to Atomism, the differences between A and K are to be attributed to different distributions of the same atoms within an unchanging void. Classical Mechanics accentuates the parity between A and K by showing that the process leading from A to K is reversible. A reversal of all the velocities in K would lead back to A.

---

[1] For the fruitfulness of these equivalent reformulations see Feynman [1965], p. 168.

Every reduction in the number of different atoms, every theory proposing say that atoms differ only in mass and charge, is hailed by science as a great step forward. This is because each such step involves a reduction of diversity, *i.e.* an increase in homogeneity. In the same spirit Einstein proposed to define simplicity in terms of the paucity of the primitive elements occurring in a given theory. In Meyersonian language, the number of primitives is a measure of diversity. Already at this stage, we recognise the close affinity which exists between Meyerson's principle of identity on the one hand and, on the other, Einstein's unity principle which provided the heuristic of Unified Field Theories.

---

[1] See Meyerson [1908], chapter 1. Also [1921], chapter 5.

The progress of science is seen by Meyerson as a struggle between the mind which tries to impose identity and a differentiated reality which resists such an imposition. The mind cannot completely succeed in denying all diversity; but some aspects of reality allow of subsumption under the principle of identity; *which* particular aspects can be determined only *a posteriori*, *e.g.* by trial-and-error methods. The general identity principle however is *a priori*.

(*i*) *Legal Explanations*

The identity principle applies at two distinct levels: at the second-order level of laws and at the first-order level of things. This gives rise to what Meyerson calls legal explanations and causal explanations respectively. Legality means that the *same* law applies irrespective of place and time, *i.e.* the form of the law is both a-spatial and a-temporal. There is thus permanence or constancy of form: many diverse phenomena illustrate, or manifest, the same underlying law which constitutes their Platonic Form.[1] For example, the falling stone, the moon revolving round the earth and the passenger thrown forward in a decelerating train are all instantiations of one and the same gravitational law. This permanence of the rule of law enables us to predict, hence has high survival value. According to Meyerson, every animal species needs for its survival this ability to anticipate facts. Thus he gives an evolutionary reason for the *a priori* character of the principle of identity as applied to laws. This principle is *a priori* in the sense of being inborn in a species which needs it for survival.

(*ii*) *Causal Explanations and Conservation Laws*

The human mind goes much further than asserting just the permanence of natural laws; it searches for causal explanations, which Meyerson defines as those under which certain substances, or things, are conserved; *i.e.* the total quantity of these substances remains unchanged through time. Conservation laws thus flow from the principle of identity *qua* negation of the diversity of things in time. I have said that Meyerson calls this version of the principle the causal one: the effect is equal to, *i.e.* it contains as much substance as, its cause.

---

[1] See Popper [1972], p. 191.

Meyerson doubts that the causal principle, as distinct from the merely legal one, has any survival value. He thinks that phenomenological laws are all that is needed for a species to gain control over nature. The causal tendency is a luxury which cannot be explained in biological terms. It can however be argued that the postulation of substances which perdure in time enables us to keep track of things, hence predict the sensations which these things supposedly cause in us. To assume that the tree exists even when I do not look at it enables me to explain, and thus predict, what I shall perceive when I turn round and look at it again. Mach claimed that the concept of material object can be replaced by that of constant functional relationships between elements of sensation; but even he admitted that concepts like that of an electron having constant mass and charge effect great economy of thought and thus have great pragmatic value. Moreover, the notion of constant functional relationship, even if it could replace that of substance, is a cumbersome notion introduced merely for reasons of philosophical purism; it is far removed both from commonsense and from scientific realism which remain our best guides for action. Thus both the legal and the causal versions of the identity principle have great survival value.

Meyerson describes the way in which we construct conservation laws as follows. The mind fixes on certain processes in nature which it hypostasises; *i.e.* it turns these *processes* into *substances* or *things* whose total quantity it then assumes to be constant. For example, commonsense anchors sense-data in material objects which are supposed to perdure in time. This commonsense hypothesis forms the basis of the classical law of the conservation of mass. Another example is that of the principle of inertia.[1] For Aristotle motion was change and velocity a ratio describing this process of change. Multiplying this ratio by mass, Descartes turned it into a *thing*, which he called quantity of motion, and whose sum total in the world remains constant. This constancy was derived from God's immutability which is another name for the principle of identity. Similarly, energy is hypostasised motion and position within a field of force (kinetic and potential energies). This automatically leads to a new conservation law, the conservation of energy law. Note that in classical physics matter is the primary substance, while momentum and energy derive their existence from matter (and charge) in motion. Special Relativity puts matter and energy on a par; so we can suppose that there exists a unique substance which our senses apprehend sometimes as radiant energy say and sometimes as hard impenetrable matter. Special Relativity thus enhances the status of energy by making it a constituent part of physical reality. This establishes a link between the Special Theory on the one hand and, on the other, General Relativity where energy considerations played an important part in the discovery of the correct field equations. In connection with Relativity, however, we should sound a note of warning: substances are generally described not by single quantities but by tensors; the fact that tensor equations are frame-independent reflects the absolute character of substances; conservation is expressed by the vanishing of the tensor's divergence.

---

[1] See Meyerson [1908], chapter 3.

Summing up, we can say that the method of applying Meyerson's identity principle to evolution in time is to single out some process common to the antecedent A and to the consequent K; then to hypostatise this process, calling it say $\phi(A)$ and $\phi(K)$; finally to assert that $\phi(A)=\phi(K)$, *i.e.* to assert a conservation law. $\phi(A)$ may for example stand for the total mass in A, or the total momentum or energy. We have seen that the human mind has a tendency to deny all forms of change; if it could have its own way, it would affirm the identity of A and K and thus abolish time. (Note that, for Meyerson, time is measured by change.) Since this proves impossible, the mind resorts to the next best thing, namely the assertion that A and K are equivalent under certain aspects. The relation $\phi(A) = \phi(K)$ is obviously an equivalence relation.

This method of hypostasising processes is also used in mathematics. For a long time functions were taken to be processes in which the value of an independent variable x gave rise to the corresponding value of $y=f(x)$. This process is clearly referred to by the word 'transformation' which is often used synonymously with 'function'. By definining function as a class of n-tuples, set theory reified this process into a thing. Hermann Weyl wrote: 'Thus we are able to subordinate genetic construction to the static existence of relations'[1]. It is obvious that the a-temporal character of mathematics makes it ideally suited to the expression of theories based on the identity principle.

(*iii*) *The Geometrisation of Nature*

So far we have considered the identity principle as applied to change in time, *i.e.* to temporal diversity. Let us now examine its application to diversity in space, or more generally to diversity in some domain of simultaneous events. Consider for example the simultaneous perception of two qualities. Two so-called secondary qualities like colour and sound strike us as being irreducibly different in a way that two frequencies are not. There is continuity between two waves giving rise to two distinct sensations like those of colour and sound, whereas the sensations themselves seem to be separated by an unbridgeable gap. This is why the mind, which is determined to eliminate diversity, tries to reduce all secondary qualities to the primary ones. This however proves insufficient. Only spatio-temporal properties are deemed completely intelligible. Mass and charge are felt to be occult, or rather heterogeneous relatively to space and time. Hence the Cartesian attempt to reduce matter to extension and explain all physical phenomena in terms of figure and movement, *i.e.* in kinematical terms. As is well-known, this Cartesian attempt failed. It was supplanted by the Newtonian system which postulates space, time, mass and force as the primitive concepts. With the emergence of electromagnetism, charge and ether were added to this list.

---

[1] Weyl [1949], p. 5.

The rise and development of the Relativity Programme can be explained in Meyersonian terms as follows. General Relativity and the sequence of Unified Field theories are the realisation of the old Cartesian ideal. According to this view, the Cartesian enterprise failed not because it was wrong-headed but because the underlying geometry was too simple, too poor, to account for the diversity of the observed phenomena. By complicating their geometry, Einstein, Weyl, Eddington and Schrödinger proposed to construct global explanations of all physical phenomena. Meyerson's philosophical framework provides an explanation of why mathematics played such an important role in the creation of the new systems, and why the Relativists tried time and again to force physics into one geometry; they were trying to *dissolve* the qualitative heterogeneity or diversity of the phenomena into the homogeneity of one unified geometry (Meyerson [1924], ch. 20). This homogeneity does not of course mean that the metric or the affine connections of the geometry do not change from one point to the next; but this is the sort of continuous numerical change which offends least against the mind's dislike of diversity. Anyway, Hermann Weyl minimises the significance of this dependence of the metric on each point of space-time in the following terms:

> The nature of the metric is *one* and is absolutely given; only the mutual orientation in the various points is capable of continuous changes and dependent upon matter. Euclidean space may be compared to a crystal, built up of uniform unchangeable atoms in the regular and rigid unchangeable arrangement of a lattice; Riemannian space to a liquid, consisting of the same indiscernible unchangeable atoms, whose arrangement and orientation, however, are mobile and yielding to forces acting upon them (Weyl [1949], pp. 87–88).

Mach's ([1909]) definition of mass as negative inverse ratio of acceleration can be seen as a first step towards the geometrisation of physics. Einstein's explanation of gravity and inertia in Riemannian terms was of course the major success of the Relativity Programme. The field equations of *GTR*, namely $R_{mn} - \frac{1}{2}g_{mn} . R = -kT_{mn}$ can be interpreted in two very different ways. Reading them from right to left we can say that energy determines, or at least modifies, geometry. However, we can also read the equations from left to right and say that, once the geometry is given, *i.e.* once the g_{mn}'s are given, then the energy content of space is *defined*. This is one way of eliminating the right hand side of the equations, *i.e.* the qualitative side, with which Einstein himself was dissatisfied. He wrote:

> By this formulation one reduces the whole mechanics of gravitation to the solution of a single system of covariant partial differential equations. The theory avoids all internal discrepancies which we have charged against the basis of classical mechanics. But, it is similar to a building one wing of which is made of fine marble (left part of the equation), but the other wing of which is built of low grade wood (right side of the equation). The phenomenological representation of matter is, in fact, only a crude substitute for a representation which would correspond to all known properties of matter ([1950], p. 81).

Einstein was clearly giving the geometrical wing precedence over the material one. In the same spirit Eddington had written:

> According to the new point of view Einstein's law of gravitation does not impose any limitation on the basal structure of the world. $G_{\mu\nu}$ may vanish or it may not. If it vanishes we say that space is empty; if it does not we say that momentum or energy is present; and our practical test whether space is occupied or not—whether momentum and energy exist there—is the test whether $G_{\mu\nu}$ exists or not (Eddington [1924], p. 120).

That this is not an idiosyncratic view of Eddington's is shown by the following quotation from Schrödinger's 'Space-Time Structure':

> I would rather you did not regard these equations [that is—$(R_{ik} - \frac{1}{2} g_{ik} . R) = T_{ik}$] as field equations, but as a definition of T_{ik}, the matter tensor. Just in the same way as Laplace's equation *div* $E = \rho$ (or $\nabla^2 V = -4\pi\rho$) says nothing but: wherever the divergence of E is not zero we say there is a charge and call *div* E the density of charge. Charge does not *cause* the electric vector to have a non-vanishing divergence, it *is* this non-vanishing divergence. In the same way matter does not *cause* the geometrical quantity which forms the first member of the above equation to be different from zero, it *is* this non-vanishing tensor, it is described *by* it ([1950], p. 99).

The Relativists tried to subject electromagnetism to the treatment which they had meted out to gravitation, *i.e.* they tried to imbed the electromagnetic field into geometry. Maxwell himself had looked upon the charge density ρ as the divergence of the electric field E; and we have just seen that this was also Schrödinger's view. By trying to derive the electromagnetic field from a geometry which would also explain gravitation, the Relativists were, in Meyersonian terms, attempting to dissolve the specificity of electrical phenomena into the homogeneity of the same underlying kinematics. Matter and charge, which had appeared to us as possessing irreducibly different qualities, would now be fused into one global geometry. Putting it in theological language: God created *one* geometry, whose various aspects are apprehended by our *senses* in different ways. This is nothing but the culmination of the Cartesian programme. The Meyersonian scheme can be summed up as follows:

Principle of Identity:

(*A*) *Identity of Form:* Legal Explanations (constancy of functional relationships, phenomenological laws).

(*B*) *Identity of Substance:* Causal Explanations.

  (*B*1) *Identity of Substance in Time:* Classical Conservation Laws (Matter, Momentum, Energy).

  (*B*2) *Identity of Substance in Space:* Unity of Matter (Cartesian Reduction of Matter to Space).

  (*B*3) *Identity in Space-Time:* Geometrisation of Nature (General Relativity, Unified Field Theories, Block Universe).

Whatever one might otherwise think of it, Meyerson's philosophy explains why scientists tried to force their physical principles into a geometrical mould and consequently injected into these principles some of the surplus structure of the mould. (See 1 (*i*) *above.*) Before leaving Meyerson one final question should be asked: is it in principle possible for us to explain the whole of reality in geometrical terms? Meyerson admits that it is not. First the geometry itself, if it is to *define* matter, charge and energy, must contain arbitrary elements. If one adopts Eddington's or Schrödinger's approach, then one can give no reason why the metric g_{mn} or the affinity Γ^l_{mn} should assume any specific values at specific points; it is no longer possible to say it is because of the presence of matter or of charge, for this very presence is to be *defined* in terms of the g_{mn}'s or the Γ^l_{mn}'s. One can for example give no reason why space should be curved and not flat. Thus the values of the metric or of the affinity are contingently given, they are simply posited. Secondly: although we have banished all qualities from physics, we have not thereby eliminated them but pushed them into a no man's land between physics on the one hand and physiology or psychology (or both) on the other. We have simply shifted the burden onto the so-called body-mind problem. In Aristotle's philosophy we could make sense of why certain bodies are *perceived* as, say, hot; according to Aristotle, this is because they *are* hot. In Relativity we have to *accept* that certain geometrical properties of the physical world surface as sensations at the level of consciousness. According to Meyerson, the transition from kinematics to sensation constitutes an irrational jump which will always remain incomprehensible to us. Moreover, physics does not completely succeed in banning all qualities from its domain: in testing physical hypotheses, we have to interpret certain propositions in operationalist terms. Hence there exist bridging principles between certain physical statements on the one hand and sense-data on the other. Finally, in its attempt to establish reversibility, and thus a kind of parity between past and future, physics has utterly failed. The entropy law underlines the fundamental asymmetry between past and future. Thus, according to Meyerson, nature sets definite limits to the applicability of the identity principle, *i.e.* to the principle of the negation of diversity, which remains an *a priori* tendency of the human mind. If one accepts that the identity principle has essential survival value, then nature has to comply with it only to the minimum extent that would enable the species to survive.

3 THE CONTINUITY BETWEEN STR AND GTR

Let us now return to the question posed at the beginning of this paper, namely: is there any continuity between *STR* and *GTR*? Is there a genuine relativistic programme, or is the word 'relativistic' there simply because various hypotheses were put forward by the same man, while the hypotheses themselves are in fact heterogeneous?

(*i*) *The Heuristic Role of the Covariance Principle*

One simplistic answer to the above question would be that the Relativity Principle (that all frames of reference belonging to some class *K* are physically equivalent) provides a continuous thread running through all the stages of the programme. The General Theory simply extends this principle from the set of inertial frames to that of all possible systems—accelerated as well as unaccelerated. The programme would thus be identified by a hard core, by one assumption shared by all the scientists working in it. Although the Relativity Principle as it stands is both vague and metaphysical, it is well-known that it played an important regulative role in the Special-relativistic modification of the laws of mechanics. However, when interpreted to mean only that all the laws of physics should be covariant with respect to some group of transformations, the principle turns out to be almost tautological. In 1917 Kretschmann pointed out that all physical laws can be covariantly rewritten and Einstein accepted this criticism as valid.

It looks therefore as if the Relativity Programme had lost its most distinctive feature since any hypothesis whatever can now be regarded as a relativistic theory. This first attempt to answer our question therefore seems to end in disaster. In his reply to Kretschmann Einstein asserted that the covariance principle should be treated as a heuristic device rather than as an ordinary statement. Of all the covariantly formulated laws, we pick out the simplest one as the most suitable. The important point to note is that a theory which at first sight looks simple may, on covariant reformulation, turn out to be very complicated. Einstein wrote:

> although it is true that one can put every empirical law in a generally covariant form, yet principle (*a*) [the principle of relativity] possesses a great heuristic power which has already brilliantly proved its mettle in the case of gravitation and which is based on the following. Of two theoretical systems, both of which are in agreement with experience that one is to be preferred which, from the point of view of the absolute differential calculus, is the simpler [*einfachere*] and the more transparent [*durchsichtigere*] one. Let one express Newtonian gravitational mechanics in the form of generally covariant equations (four-dimensionally) and one will surely be convinced that principle (*a*) excludes this theory from the practical if not from the theoretical point of view (Einstein [1918])[1].

It is obvious from the above passage that Einstein meant 'simplicity' and 'clarity' in a pragmatic sense. Although words can be redefined in an arbitrary way, 'simplicity' is to my mind a misleading term, which ought to be replaced by something like 'compactness' or 'organic unity'. Even when covariantly formulated, Newton's theory cannot be said to be much more complicated than Einstein's; its main disadvantage is that it contains two heterogeneous parts: the degenerate metric and the gravitational field. The latter is described by an affine connection not determined by the metric alone, so that we have to go beyond the metrical aspect of geometry in order to capture the field. In Einstein's theory the metric tensor and the gravitational potential are one and the same thing, the affinity $-\Gamma^{\lambda}_{\mu\nu}$ and the Christoffel quantities $\{^{\lambda}_{\mu\nu}\}$ are identical (in John Stachel's words, there is 'minimal coupling' between the field and the geometry). This is what is meant by saying that Einstein tried to geometrise physics, or equivalently, to make geometry empirical.[2]

We now return to the question concerning the close connection between Gravitation and General Relativity. In dealing with this question we shall discover further bridges leading from the Special to the General Theory and thus be able to solve the continuity problem more satisfactorily. The analysis will underline the fundamental role played by the equation '$E = mc^2$' at two crucial points in the development of the programme; it will explain the heuristic function both of the principle of equivalence and of Poisson's classical equation[3] $\nabla^2\phi = 4\pi G\rho$ regarded as a limiting case of the new theory; it will also vindicate my claim that mathematics, taken as a heuristic instrument, can lead to the discovery of new physical laws. Finally, the problem as to whether the Relativity programme possesses any hard core at all will be reconsidered and a close parallel established between the positive heuristics of the Special and of the General theories.

---

[1] See Appendix.

[2] See above section 1 (*i*).

[3] The field is $-\nabla\phi$, G is Newton's gravitational constant.

(*ii*) *The Abandonment of Special Relativity*

In his [1916] paper Einstein explains his reasons for abandoning the Special Theory and for recognising the close connection between Gravitation and General Relativity. His argument is as follows. Both classical and special relativistic physics give a privileged status to an infinite set of inertial frames without relating these frames to the distribution of matter in the universe. Such a distinction between inertial and non-inertial frames is illegitimate from a Machian viewpoint (see my [1977]). Einstein therefore proposes to put all frames on equal footing, so that the following situation arises. Let K be any frame with respect to which a particle left to itself either remains at rest or goes on moving uniformly along a straight line. Consider another frame K' which is accelerating relatively to K. As viewed from K' the particle P experiences an acceleration $\mathbf{a}$ which, being due to the motion of K', must be independent of the internal composition of P. Since we have agreed to look upon K' as a legitimate frame of reference, we must regard P as acted upon by a real physical field $\mathbf{I}$ which exerts on P a force $m\mathbf{a}$ proportional to the inertial mass m of P. In classical physics we are acquainted with one field to which every material body responds proportionately to its mass, namely the gravitational field $\mathbf{G}$. Einstein assumes that $\mathbf{I}$ is of the same nature as $\mathbf{G}$, so that any general theory of relativity will be closely bound up with gravitational phenomena.

Because of Einstein's references to Mach, the myth arose that an *idealist* philosophy, namely Machism, was the starting point of a revolutionary *physical* theory. We have however seen that the covariance postulate is *vacuous*, so it certainly makes no philosophical presuppositions, whether idealist or realist.[1] Even when interpreted in a prescriptive heuristic sense, the covariance principle has to be supplemented by a 'compactness' requirement in order to become effective. The nearest analogue to compactness in Machist philosophy is the economy of thought principle, according to which scientific theories are convenient devices for subsuming an infinite number of facts under a limited number of laws. Thus 'economy of thought' is more akin to 'simplicity' than to 'unity' or 'compactness'; and a theory may be more unified and yet less simple than another. There is of course no reason to doubt that Einstein was influenced by Mach, who, in his 'Conservation of Energy', came very near to formulating a covariance principle. However, this influence has nothing to do with Machism as such but only with Mach, and also it is not, as will presently be shown, the only reason why Einstein abandoned Special Relativity. (See my [1977].)

---

[1] See however Appendix.

After deriving the Lorentz transformation Einstein, and after him Planck, modified Newton's second law so as to make it Lorentz-covariant. It was to be expected that Einstein would try a similar procedure with the gravitational equations, notwithstanding his Machian qualms about the absolute nature of the inertial frames. In fact this is precisely what he attempted to do up to 1908 when, in view of the equivalence principle, he decided that the undertaking was hopeless.

In the case of Special Relativity Einstein's starting point was a 'fact' discovered by Faraday; the fact namely that, in electromagnetic induction, the 'observed' result depends on the relative velocity of the conductor and magnet and not on their absolute motion in the ether. (See my [1973].) In the case of General Relativity the result from which Einstein started goes back even further, in fact it goes back to the early seventeenth, or late sixteenth, century. Galileo discovered that all bodies near the surface of the earth fall freely with same acceleration. A two-fold constancy is involved here: the same body B has a constant acceleration throughout its time of fall; and all bodies irrespective of their mass and composition experience the same acceleration at the same point of the trajectory. The first type of constancy was rejected by Newton, but the second one can be deduced from his theory as follows.

A body P has three different masses:

(1) Inertial mass m which measures the capacity of the body for resisting the action of applied force;

(2) Active gravitational mass m' through which it contributes to the gravitational field; and finally

(3) Passive gravitational mass m'' by which it responds to an existing gravitational field.

Given two mass points P_1 and P_2, the force acting on P_2 due to the presence of P_1 is $\frac{m_1' \, m_2''}{r^2}\,\mathbf{u}$. Similarly the force on P_1 is $\frac{m_1'' \, m_2'}{r^2}(-\mathbf{u})$. By Newton's third law: $\frac{m_1' \, m_2''}{r^2}\,\mathbf{u} = -\frac{m_1'' \, m_2'}{r^2}(-\mathbf{u})$. Hence $m_1' m_2'' = m_1'' m_2'$ and so $m_1'/m_1'' = m_2'/m_2''$.

We can choose our units so as to make $m_1' = m_2''$ and $m_1' = m_2''$. The two gravitational masses are therefore equal. By the second law of motion: $\frac{m_1' m_2'}{r^2}\mathbf{u} = m_2 \mathbf{a}_2$, where $\mathbf{a}$, is the acceleration of P_2. At this point Newton blandly assumes that the gravitational and inertial masses are equal, so that $m_2 = m_2'$ and $m_1 = m_1'$; hence $\frac{m_1 \mathbf{u}}{r^2} = \mathbf{a}_2$, in view of $\frac{m_1 m_2}{r^2}\mathbf{u} = m_2 \mathbf{a}_2$. It follows that, whatever its material composition, P_2 experiences the same acceleration $\mathbf{a}_2$, provided that m_2 is small enough not to affect the motion of P_1 in any appreciable way.

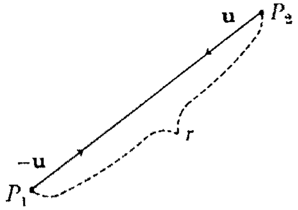


Figure 1

The gravitational field is therefore a kinematical, or more precisely an acceleration, field. It is characterised by the direction and magnitude at each point of the acceleration which any test body experiences, irrespective of its composition and velocity. We can neutralise the field at each point by choosing a frame whose acceleration is equal to the field, in other words by letting the coordinate system fall freely with the field.

Hence we have a 'fact', namely the equality of gravitational and inertial masses, which is unsatisfactorily accounted for by the classical theory. We shall see how General Relativity remedies this defect, but before pursuing this line of thought any further, let us examine what conditions a Special Relativistic theory would have to meet in order to account for gravitational phenomena. In a reply to Max Abraham, Einstein wrote:

> One of the most important results of Relativity Theory is the knowledge that every form of energy E possesses inertia E/c^2 proportional to E. Since, as far as our experience goes, every inertial mass is at the same time a gravitational mass, we cannot but attribute to every form of energy E a gravitational mass equal to E/c^2. From this it immediately follows that the gravitational force acting on a body is greater when the body is in motion than when the body is at rest.
>
> If the gravitational field could be accounted for in present day Relativity Theory, this could only happen in one of the following ways. One can regard the gravitational field either as a 4-vector or as an antisymmetric tensor of the second order. . . . One thereby obtains results which contradict the above-mentioned consequences concerning the gravitational mass of energy. It there-

fore looks as if the gravitational vector cannot be consistently fitted into the Relativistic scheme as it stands at the present moment (Einstein [1912]; my translation).

What follows is a technical illustration of the point made by Einstein in the above passage. Let **f** be the gravitational 3-field in some frame of reference. Consider a body P of rest mass m_0 moving under the effect of the gravitational force. Since $\frac{m_0}{\sqrt{1-v^2/c^2}}$ is the inertial mass of the body, it is also, according to Einstein, the gravitational mass. In other words $\frac{m_0}{\sqrt{1-v^2/c^2}}$ is the measure of P's response to the field. By the equation of motion we have:

$$\frac{d}{dt}\left(\frac{m_0\mathbf{v}}{\sqrt{1-v^2/c^2}}\right)=\frac{m_0}{\sqrt{1-v^2/c^2}}\,\mathbf{f}$$

Through expanding the left-hand side of this equation, then dividing by $m_0/\sqrt{1-v^2/c^2}$, we obtain:

$$\frac{1}{m_0}\cdot\frac{dm_0}{dt}\,\mathbf{v}+\mathbf{a}+\sqrt{1-v^2/c^2}\,\frac{d}{dt}\,[1/\sqrt{1-v^2/c^2}]\,\mathbf{v}=\mathbf{f}$$

where **a** is the acceleration of P. If the body is heated and so gains energy during the motion, then, in view of $E=mc^2$, m_0 varies with time. Hence $dm_0/dt\neq 0$, so **a** depends on the internal composition of P, which contradicts the principle of equivalence. Even if the rest mass remains constant, **f** still depends on **v** and has thus ceased to be a pure acceleration field.[1]

The situation is no better if the field is described by an antisymmetric tensor H_{uv}. In the electromagnetic case the 4-force acting on a charged particle is given by $ke\ F^{\mu}\sigma V^{\sigma}$, where k is a constant, e is the charge and $F^{\mu\sigma}$ the electromagnetic tensor. Hence $d/ds\,(m_0V^{\mu})=keF^{\mu}\sigma V^{\sigma}$.
In the case of gravitation the equation would be as follows:

$$\frac{d}{ds}(m_0\,V^{\sigma})=km_0\,H^{\sigma}_{\mu}\,V^{\mu}\,;\quad i.e.\ \frac{1}{m_0}\,\frac{dm_0}{ds}\,V^{\sigma}+\frac{dV^{\sigma}}{ds}=kH^{\sigma}_{\mu}\,V^{\mu}$$

Multiplying by V_{σ} and noting that $V_{\sigma}\,V^{\sigma}=1$, hence that $V_{\sigma}\,\frac{dV^{\sigma}}{ds}=0$, we obtain $\frac{1}{m_0}\cdot\frac{dm_0}{ds}=0$. Thus $dm_0/ds=0$, which means that m_0 must remain constant. This contradicts Special Relativity.

In sum, Einstein had the following two reasons for giving up Special Relativity as a suitable framework for the whole of physics:

---

[1] For this second point, namely that **f** has anyway ceased to be a pure acceleration field, I am indebted to Professor J. Stachel.

P2223

(1) Philosophical dissatisfaction with having given a privileged status to the set of all inertial frames of reference.

(2) The technical difficulties, in view of $E = mc^2$, of accommodating Gravitational Theory with Special Relativity.

The second reason appears to have been the more decisive one.

The process through which the Special Theory was abandoned can be described as follows. It was precisely the law concerning the interchangeability of mass and energy, one main result of Special Relativity, which brought about the downfall of the latter. Special Relativity as it were succumbed to its inner 'contradictions', more particularly to the impossibility, in Einstein's view, of reconciling $E = mc^2$ with gravitation.

(*iii*) *The Heuristic Role of the Principle of Equivalence*

In his [1912] Einstein wrote:

> In the immediate future our task will be to create a relativistic scheme in which to express the equivalence of gravitational and inertial masses. I made a first, albeit modest, attempt towards achieving this aim in my own works on the static gravitational field. In these I started from the obvious consideration that the equivalence of gravitational and inertial masses is to be explained in terms of the essential identity (*Wesensgleichheit*) of these two elementary qualities of matter (*viz*. energy) as follows: a static gravitational field is to be regarded from the physical point of view as essentially identical (*Wesensgleich*) with an acceleration of the frame of reference. I have to confess that I could consistently carry through this interpretation only for infinitely small regions, and that I can give no satisfactory reasons for this state of affairs. (My translation.)

The most striking word in this passage is '*Wesensgleichheit*' or 'identity of essence', which reveals an essentialist streak in Einstein's scientific thinking. A method he often used was to question well-known 'facts' and isolate certain features which are unsatisfactorily explained by extant hypotheses. Newtonian theory accounts for the 'fact' that all bodies fall with the same acceleration by postulating the equality of gravitational and inertial masses ($m = m'$). For Einstein this equality is a manifestation of a deep *essential* identity. Bodies fall with the same acceleration because gravitational fields are so to speak created by local accelerations of the frame of reference.

The last two sentences in the above passage indicate what difficulties Einstein was facing in 1912. He could explain the behaviour of the field in infinitely small regions without however being able to construct a global theory of gravitation.

Consider a point P at which the field is represented by the acceleration $\mathbf{g}$. In an infinitely small neighbourhood N of P the field can be regarded as constant.

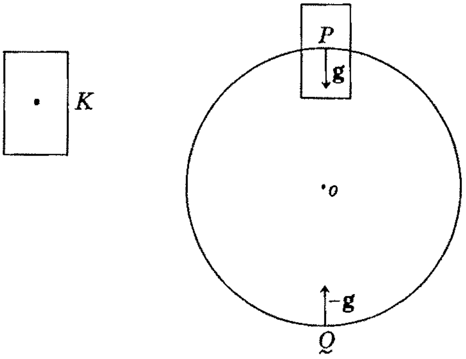


Figure 2

Let K be our original frame and K' the system of reference whose acceleration with respect to K is $\mathbf{g}$. The equivalence principle states that K' can be considered as field-free[1] and the field in K as generated by the acceleration of K relatively to K'. However, this argument only applies to the immediate neighbourhood of P; if for example the field is created by a single particle at O and Q is the point diametrically opposed to P with respect to O, the field in K' at Q is $-2\mathbf{g}$. K' is inertial or field-free only at P; in fact transforming the field away at P entails piling it up somewhere else (at Q for example).

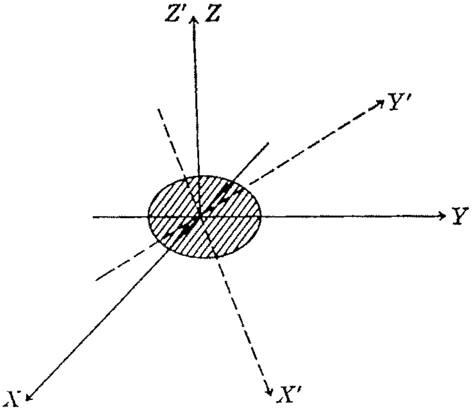


Figure 3

Thus, Einstein found himself compelled to abandon Euclidean kinematics. An important thought-experiment which finally convinced him that Euclidean geometry had to be given up is the so-called rotating disc experiment.[2] Starting from Special Relativity let us consider an inertial frame OXYZ with respect to which a material disc rotates in the XOY plane. It is important to look upon the disc as already rotating and not to wonder how its motion arose in the first place. Let an observer who moves with the disc, *i.e.* who is at rest relatively to it, measure its circumference and its radius and then divide the results of these two measurements. He will obtain, not 2π, but $2\pi/\sqrt{1-w^2r^2/c^2}$ where w is the angular velocity of the disc; for, by the Lorentz contraction, the rods he places along the circumference will be shortened by the factors $\sqrt{1-w^2r^2/c^2}$, while the rods he uses for measuring the radius, being transverse to the motion, are unaffected by it. Since the result $2\pi/\sqrt{1-w^2r^2/c^2}$ is independent of the composition both of the rod and of the disc, the observer will conclude that his (2-dimensional) geometry is non-Euclidean. Einstein thus faced the problem of explaining gravity locally in terms of some non-Euclidean geometry. In 1913–14 he found a solution to this problem by turning the principle of equivalence into a powerful heuristic instrument. He looked at it from the following new angle.

---

[1] We shall see that this is not strictly speaking true.

[2] *Cf.* Stachel [1979].

Consider a classical inertial frame K and the Newtonian gravitational field created by a single mass point at O. Let K' be an accelerated frame with the same origin O, and P a material point of (negligible) gravitational mass m' and inertial mass m. By the second law of motion the total force $\mathbf{F}$ acting on P is given by :

$$\mathbf{F} = m\frac{d^2}{dt^2}(\xi_i\mathbf{w}_i) = m(\ddot{\xi}_i\mathbf{w}_i + 2\dot{\xi}_i\dot{\mathbf{w}}_i + \xi_i\ddot{\mathbf{w}}_i)$$

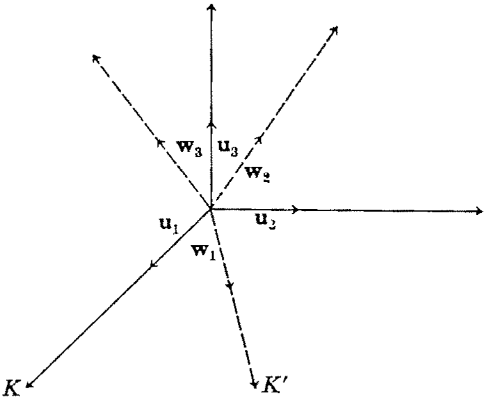


Figure 4

where (ξ_i) are the components of OP in K' and the dots represent differentiations with respect to time. $\ddot{\xi}_i\mathbf{w}_i$ is therefore the acceleration α of P in K'. Putting $\mathbf{I} = -(2\dot{\xi}_i\dot{\mathbf{w}}_i+\xi_i\ddot{\mathbf{w}}_i)$, we obtain $\mathbf{F} = m\alpha - m\mathbf{I}$, hence: $m'\mathbf{g}+\mathbf{H} = m\alpha - m\mathbf{I}$, where $\mathbf{g}$ is the gravitational field at P and $\mathbf{H}$ is the resultant of the external forces exerted on P. Since $m = m'$, the last equation can be rewritten as:

$$\mathbf{H}+m(\mathbf{g}+\mathbf{I}) = m\boldsymbol{\alpha}. \tag{1}$$

Note that **I**, the inertial force, is independent of the composition of P. Everything happens in K' as if an extra inertial field **I** had been added, where **I**, like Newtonian gravitation, acts on P proportionately to its mass (inertial or gravitational). In classical physics we always regard $m\mathbf{I}$ as an artificial force which can be transformed away by choosing K' to be an inertial frame, for example by making K' coincide with K. This option, however, is no longer open to a physicist who has given up the idea of an *a priori* privileged set of inertial systems. He has now to look upon K' as a legitimate frame and upon **I** as a real physical field. In K' however the fields **g** and **I** cannot be experimentally told apart in the following sense. By equation (1):

$$\mathbf{g}+\mathbf{I} = \alpha - 1/m\,\mathbf{H}.$$

In whatever way we vary the test-particle at P, we shall be able to solve this equation only for the sum $(\mathbf{g}+\mathbf{I})$ and not for **g** and **I** separately. If, instead of a mass at O, we had an electric charge creating an electrostatic field **E**, then equation (i) would become: $\mathbf{H}+e\mathbf{E}+m\mathbf{I} = m\boldsymbol{\alpha}$, where e is the charge carried by P. Hence:

$$\frac{e}{m}\mathbf{E}+\mathbf{I} = \alpha - \frac{1}{m}\mathbf{H}. \tag{2}$$

By varying e/m through taking different test-particles, we can solve two equations of type (2) in the two unknowns **E** and **I**. Thus **E** and **I** can be told apart in a way in which **g** and **I** cannot.

Since gravitational and inertial fields always occur indissolubly mixed, Einstein typically refused to consider this 'fact' a mere accident. The division between inertial and strictly gravitational components is artificial; there exists only one total G-field, namely $\mathbf{G} = \mathbf{g}+\mathbf{I}$, which a new theory would have to account for. In view of $\mathbf{G} = \mathbf{g}+\mathbf{I}$, **I** is now a special case of a G-field, namely that variety which can be completely transformed away by a single global change of coordinates. In the above example this would correspond to the case where the mass at O vanishes and the field in K' was entirely due to the acceleration of the frame; the field vanishes when we go over from K' to K.

By a stroke of genius Einstein put this result to heuristic use as follows. Reducible G-fields can be generated at will by accelerating frames of references relatively to those systems in Minkowski space in which both the equation

$$ds^2 = (dx^0)^2-(dx^1)^2-(dx^2)^2-(dx^3)^2 = c^2dt^2-dx^2-dy^2-dz^2$$

and the law of inertia hold. In the accelerating frame K', ds^2 will assume the form: $ds^2 = g_{\mu\nu}dx^{\mu}dx^{\nu}$, where the matrix $(g_{\mu\nu})$ is reducible to

$$\begin{pmatrix} 1 & & & \\ & -1 & & \bigcirc \\ \bigcirc & & -1 & \\ & & & -1 \end{pmatrix}$$

by a global transformation of the coordinates. In K' we can study the behaviour of the G-field generated by the acceleration of the frame then extend our results to the irreducible case by lifting the restrictions initially imposed on the $g_{\mu\nu}$'s [namely the restriction that $(g_{\mu\nu})$ should be reducible to a constant matrix by a single global transformation]. In order to carry out such a programme however, *our results in* K′ *have to be expressed in terms of the* $g_{\mu\nu}$*'s directly, without reference to any special properties which the latter quantities may initially have possessed. This is where mathematics and more particularly the absolute differential calculus played a crucially important role as will presently be shown.*

Using this method, Einstein determined the path followed by a particle falling freely in a gravitational field. In a frame K where no G-field is present and $ds^2 = (dx^0)^2-(dx^1)^2-(dx^2)^2-(dx^3)^2$ the trajectory of a free particle is a straight line whose equations are:

$$d^2x^i/(dx^0)^2 = 0 \quad \text{for} \quad i = 1, 2, 3.$$

Since our aim is to look at the same particle from an accelerated system, it will be useful to characterise the path in an invariant way. Let us therefore use the intrinsic parameter s instead of x^0 as the independent variable. Since $d^2x^i/(dx^0)^2 = 0$ we have $dx^i/dx^0 = \text{constant}$, hence:

$$\left(\frac{dx^1}{dx^0}\right)^2+\left(\frac{dx^2}{dx^0}\right)^2+\left(\frac{dx^3}{dx^0}\right)^2 = k^2 \ (k \text{ being a constant});$$

i.e. $(dx^1)^2+(dx^2)^2+(dx^3)^2 = k^2(dx^0)^2$.

Therefore: $ds^2 = (dx^0)^2-(dx^1)^2-(dx^2)^2-(dx^3)^2 = [1-k^2](dx^0)^2$.

$$ds = h.\,dx^0, \text{ where } h = \pm\sqrt{1-k^2} = \text{constant}.$$

Since $\dfrac{d^2x^i}{(dx^0)^2} = 0$, it follows that: $\dfrac{d^2x^i}{ds^2} = 0$ for $i = 1, 2, 3$.

Moreover, $\dfrac{d^2x^0}{ds^2} = 0$, in view of $ds = h\,.\,dx^0$. Hence $\dfrac{d^2x^{\mu}}{ds^2} = 0$ for $\mu = 0, 1, 2, 3$.

(3) The equations $\dfrac{d^2x^{\mu}}{ds^2} = 0$ give a stationary value to the integral $\int ds$, because $\delta \int ds = 0$.

Since ds *is an invariant, this result applies to all frames of reference.* Therefore, in K' the path of the same particle will be given by $\delta \int ds = 0$. This means that in K', or in any other frame of reference, the path will be a geodesic, provided we take the infinitesimal distance ds to be given by:

$$ds^2 = g_{\mu\nu}\, dx^\mu dx^\nu. \tag{4}$$

From the absolute differential calculus we know that a geodesic satisfies the following set of differential equations:

$$\left\{\begin{aligned} &\frac{d^2x^\mu}{ds^2} + \left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = 0; \quad \text{where:} \\ &\left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\} = g^{\mu\sigma}[\alpha\beta, \sigma] = \tfrac{1}{2} g^{\mu\sigma}\left(\frac{\partial g_{\alpha\sigma}}{\partial x^\beta} + \frac{\partial g_{\beta\sigma}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\sigma}\right) \end{aligned}\right. \tag{5}$$

Comparing (3) and (5) we conclude that in the accelerating frame the trajectory of the particle is *no longer straight in the ordinary sense of the word* and that $\left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\}$ is the quantity which deflects the particle from a straight path. Since $\left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\}$ is a function of the $g_{\alpha\beta}$'s and since we ascribe the deviation from a straight trajectory to the action of a G-field, the latter is represented by the $g_{\alpha\beta}$'s or rather by the partial derivatives $\partial g_{\alpha\beta}/\partial x^0$; for the quantities $\left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\}$ vanish if the $g_{\alpha\beta}$'s are constant. The $g_{\alpha\beta}$'s can thus be regarded as gravitational potentials.

So far we have assumed that the G-field is reducible (inertial, in the old terminology). Applying the method described above, we now generalise our results by extending them to the case where there need not exist a global transformation making the $g_{\alpha\beta}$'s constant throughout the manifold. Thus, even if the field is irreducible, the path of a free particle will be a geodesic and the metric tensor will still represent the gravitational potential.

In this process of generalisation the importance of the absolute calculus cannot be overestimated. Right from the beginning the path was characterised in an invariant way as a curve for which $\delta \int ds = 0$. In the differential equation $\frac{d^2x^\mu}{ds^2} + \left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\}\frac{dx^\alpha}{ds}\cdot\frac{dx^\beta}{ds} = 0$, the Christoffel quantities $\left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\}$ are directly related to the $g_{\alpha\beta}$'s *without reference to any special properties which the* $g_{\alpha\beta}$*'s might have.* In other words: Einstein had at his disposal a calculus in which *the functional dependence of the affinities on the metric is exhibited independently of the frame chosen and also of the peculiar properties of the space under consideration.* Had we started from the equations $\frac{d^2x^i}{dt^2} = 0$ and tried to go over directly to an accelerated frame, it would have been virtually impossible to get away from flat space.

The same technique was applied to writing down Maxwell's equations in generally covariant form, then generalising them to the case where the G-field is irreducible. In a Lorentzian frame the equations for the electromagnetic field are:

$$(6)\quad \frac{\partial F_{\mu\nu}}{\partial x^{\sigma}}+\frac{\partial F_{\nu\sigma}}{\partial x^{\mu}}+\frac{\partial F_{\sigma\mu}}{\partial x^{\nu}}=0; \quad \textit{i.e.} \quad [F_{\mu\nu\,|\,\sigma}]=0$$

$$(7)\quad \frac{\partial F^{\mu\nu}}{\partial x^{\nu}}=S^{\mu}; \quad \textit{i.e.} \quad F^{\mu\nu}_{\ \ |\nu}=S^{\mu}$$

Equation (6) is already in generally covariant form. As for (7) we notice that, since the $g_{\mu\nu}$'s are constant, the affinities $\begin{Bmatrix}\mu\\ \alpha\beta\end{Bmatrix}$ vanish, so the ordinary and covariant derivatives coincide. Hence (7) can be written as:

$$(7')\qquad F^{\mu\nu}_{\ \ \|\nu}=S^{\mu}.$$

which is generally covariant. We now assume that, even when space is curved, (6) and (7′) still hold.

One last example illustrating the way in which the principle of equivalence can be used heuristically was given by Einstein in his [1914] paper.

If $T^{\mu\nu}$ is the total energy-tensor in flat space, then the conservation laws for energy and momentum take on the form $T^{\mu\nu}_{\ \ |\nu}=0$. This equation is equivalent to $T^{\mu\nu}_{\ \ \|\nu}=0$ because $\begin{Bmatrix}\mu\\ \alpha\beta\end{Bmatrix}=0$ for all μ, α, β. But $T^{\mu\nu}_{\ \ \|\nu}=0$ is generally co-variant, so it still holds when the frame of reference is accelerated. We now go over to the general case by postulating that $T^{\mu\nu}_{\ \ \|\nu}$ vanishes even when an irreducible field is present, *i.e.* even when the space ceases to be flat.

The question arises as to whether it is really necessary to give this heuristic twist to the principle of equivalence. Can it not be simply treated as an ordinary statement which strictly entails an extension of Special relativistic laws to cases where an irreducible G-field is present? This seems to be the position of Eddington who says that the transition from $T^{\mu\nu}_{\ \ |\nu}=0$ to $T^{\mu\nu}_{\ \ \|\nu}=0$ involves an appeal to the principle of equivalence. The reasoning runs as follows: take any law such as $T^{\mu\nu}_{\ \ |\nu}=0$, which holds in gravitation-free space. At an arbitrary point P of curved space the field can be transformed away and a geodesic frame chosen in which all the affinities vanish. Since the field has disappeared, $T^{\mu\nu}_{\ \ |\nu}=0$ holds at P and, because the affinities vanish, $T^{\mu\nu}_{\ \ |\nu}\equiv T^{\mu\nu}_{\ \ \|\nu}$; hence $T^{\mu\nu}_{\ \ \|\nu}=0$ and this equation, being generally covariant, is true in the original frame. P being an arbitrary point, $T^{\mu\nu}_{\ \ \|\nu} = 0$ holds throughout the manifold. This seems to be Eddington's argument which obviously depends on the meaning of 'transforming the field way'. We can of course choose a (geodesic) system such that $(g_{\mu\nu})$ assumes the value

$$\begin{pmatrix} 1 & & & \\ & -1 & & \text{\Large 0} \\ \text{\Large 0} & & -1 & \\ & & & -1 \end{pmatrix}$$

and $g_{\mu\nu|\sigma} = 0$ for all μ, ν, σ. The $g_{\mu\nu}$'s are thus quasi-constant in the neighbourhood of P. However, all the second derivatives of the $g_{\mu\nu}$'s cannot be made to vanish at an arbitrary point P; otherwise the curvature tensor would also vanish and, being a tensor, it will be zero in all frames of reference; the space would thus have to be flat, which is impossible in the case of an irreducible field. We cannot generally say of a geodesic frame that it is strictly Lorentzian and so conclude that some special-relativistic law must hold at P.

So sum up: by turning the equivalence principle into a heuristic instrument, Einstein solved the problem of constructing a global theory of gravitation. Two points can be made in this connection:

(1) If taken to mean that the G-field can be completely transformed away at each point, the principle of equivalence is strictly speaking false. It would, if so interpreted, only apply to flat space.

(2) However, the principle of equivalence can be heuristically exploited as follows. We start from a Lorentzian reference system, generate a reducible G-field by accelerating the frame, study the properties of such a particular field, then go over to the case of an irreducible field. *The last step however involves a genuine generalisation.*

Once again, we have a statement which, when interpreted literally, is actually false; only as part of the positive heuristic guiding the programme is its full significance realised.

(*iv*) *The Field Equations*

So far we have regarded the $g_{\mu\nu}$'s as given and determined the effects of gravitation on other physical processes such as the electromagnetic phenomena. The path of a free particle was found to be a geodesic, *i.e.* a path which 'extremises' the value of $\int ds = \int \sqrt{g_{\mu\nu}\, dx^\mu\, dx^\nu}$. The equation of a non-null geodesic is:

$$\frac{d^2x^\mu}{ds^2} + \begin{Bmatrix} \mu \\ \alpha\beta \end{Bmatrix} \frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = 0.$$

If an electromagnetic field is present, the trajectory of a charged particle deviates from a geodesic according to the following equation of motion:

$$-m_0\left(\frac{d^2x^\mu}{ds^2}+\left\{\begin{matrix}\mu\\ \alpha\beta\end{matrix}\right\}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds}\right)=k\rho_0 F^\mu{}_\nu\frac{dx^\nu}{ds},$$

(k = constant, ρ_0 = proper density of charge).[1]

Einstein gave these results at the beginning of his 1914 paper, somewhat in the same way as Newton proposed his second law of motion before proceeding to 'derive' his inverse-square law. Thus all that remained for Einstein to do was to lay down field equations for the $g_{\mu\nu}$'s, but this was precisely where most of the difficulties arose.

We have already seen how Einstein decided that the gravitational potential was to coincide with the metric; this resulted from considering the gravitational force as due to an acceleration field determined by the geometry of space-time. The principle of equivalence also tells us that the G-field in any frame of reference must be regarded as a real physical field; by Newtonian theory such a field is determined by the distribution of mass throughout space (and boundary conditions) and, by Special Relativity, mass and energy are interchangeable. It follows from the conjunction of all these statements that the gravitational potential or metric must be related to the energy content of space, which in Special Relativity is described by a symmetric tensor $T^{\mu\nu}$.

In 1913 Einstein and Grossmann[2] published a joint paper to which the former contributed the first-physical-section and the latter the mathematical section. Einstein proposed to find field equations of the form: $H^{\mu\nu} = -kT^{\mu\nu}$ where $H^{\mu\nu}$ was to be a function of the $g_{\alpha\beta}$'s and their derivatives. He was guided by the idea that Poisson's equation $\nabla^2\phi = 4\pi G\rho$ must be yielded as a limiting case of his own solution. From Poisson's equation he drew the following two conclusions, of which the second one, partly because it was not explicitly stated but taken for granted, proved an obstacle to further progress.

(a) Like $\nabla^2\phi$, $H^{\mu\nu}$ was to be of the second order of differentiation in the potentials $g_{\alpha\beta}$'s.

(b) $\nabla^2\phi = 4\pi G\rho$ is a single independent equation for a single unknown function ϕ. Einstein therefore expected his own solution to consist of ten independent equations for the ten functions $g_{\mu\nu}$; also, given suitable boundary conditions, the $g_{\mu\nu}$'s were to be uniquely determined.

We shall not go into Einstein's proposed equations, which were later modified. It is far more interesting for the understanding of subsequent developments to concentrate on Einstein's concluding remarks and on an approach which was initiated by Grossmann. Towards the end of the paper Grossmann writes that, since Riemann and Christoffel, the geometrical properties of space were known to be closely connected with the curvature tensor $B^{\alpha}{}_{.\mu\nu\sigma}$. For example the vanishing of this tensor is the necessary and sufficient condition for the flatness of space. Consequently $B^{\alpha}{}_{.\mu\nu\sigma} = 0$ is too stringent a law in the case of an irreducible field. Looking at Einstein's law $H_{\mu\nu} = -kT_{\mu\nu}$, Grossmann proposes to substitute for $H_{\mu\nu}$ a symmetric tensor $R_{\mu\nu}$ constructed from the $B^{\alpha}{}_{.\mu\nu\sigma}$'s alone. One standard way in the absolute calculus of lowering the rank of tensors is through contraction of indices. Note that $B^{\alpha}{}_{.\mu\alpha\nu} = -B^{\alpha}{}_{.\mu\nu\alpha}$, since $B^{\alpha}{}_{.\mu\nu\beta}$ is antisymmetric in ν and β. If we contract α and μ, we obtain an expression which vanishes identically: for: $B^{\alpha}{}_{.\alpha\nu\beta} = g^{\alpha\sigma}B_{\sigma\alpha\nu\beta} = g^{\sigma\alpha}B_{\sigma\alpha\nu\beta} = -g^{\sigma\alpha}B_{\alpha\sigma\nu\beta}$ (by antisymmetry of $B_{\sigma\alpha\nu\beta}$ in σ and α) $= -B^{\sigma}{}_{.\sigma\nu\beta} = -B^{\alpha}{}_{.\alpha\nu\beta}$. Consequently $B^{\alpha}{}_{.\alpha\nu\beta} = 0$. Thus there exists essentially one way of contracting two indices in $B^{\alpha}{}_{.\mu\nu\beta}$, namely the contraction of α with β. One form of the gravitational field equations which readily suggests itself is therefore:

---

[1] See Eddington [1924], § 80.

[2] Einstein and Grossmann [1913].

$$R_{\mu\nu} = -kT_{\mu\nu}, \tag{1}$$

where:

$$R_{\mu\nu} = B^{\alpha}{}_{.\mu\nu\alpha}.$$

This example illustrates the first kind of heuristic role which mathematics can play in physical discovery. (see section 1 (*i*) above). Grossmann starts from a vague physical principle (geometry is connected with the energy content of space) which, through being translated into a mathematical theory (namely, Riemannian Geometry), becomes a much stronger proposition, namely $R_{\mu\nu} = -kT_{\mu\nu}$. Grossmann however rejects this law on the grounds that it does not admit Poisson's equation as a limiting case.[1] With hindsight we realise that, although he was right in rejecting the law, Grossmann missed by little the opportunity of proposing the 'correct' field equations, namely $R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -kT_{\mu\nu}$. The irony of course is that for empty space[2]: $(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu} R = 0) \Leftrightarrow (R_{\mu\nu} = 0)$ so that $R_{\mu\nu} = 0$ holds after all in this special case.

At the end of the paper Einstein gives a reductio ad absurdum proof of the impossibility for the field to be determined by a generally covariant law. In fact he showed, without being fully aware of it, the incompatibility of the following two statements:

(1) The field equations are generally covariant.

(2) Given suitable boundary conditions, the equations completely determine the potentials.

---

[1] In a letter to Summerfeld Einstein admits that this was a (technical) mistake (see Einstein [1915*e*].

[2] Suppose $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu} R = 0$. Multiplying by $g^{\sigma\mu}$: $R^{\sigma}_{\nu} - \frac{1}{2}\delta^{\sigma}_{\nu} R = 0$. contracting σ and ν: $0 = R - \frac{4}{2}R$; *i.e.* $R = 0$. Hence, in view of $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu} R = 0$, we have: $R_{\mu\nu} = 0$.

Since he took (2) for granted, he concluded that (1) must necessarily be false. His proof is roughly as follows:

Suppose the field equations to be of the form $E_{\alpha\beta}(g_{\mu\nu}, g_{\mu\nu|\gamma}, g_{\mu\nu|\gamma\sigma}) = 0$ where the left-hand side is a tensor. If in some frame of references $(x^1, \ldots, x^N)$, $g_{\mu\nu} = \phi_{\mu\nu}(x)$ is a solution then

$$\bar{g}_{\mu\nu} = \bar{\phi}_{\mu\nu}(\bar{x}) = \frac{\partial x^\alpha}{\partial \bar{x}^\mu} \cdot \frac{\partial x^\beta}{\partial \bar{x}^\nu} \cdot \phi_{\alpha\beta}(x)$$

is a solution in any other coordinate system $(\bar{x}^1, \ldots, \bar{x}^N)$,
i.e. $[E_{\alpha\beta}(\phi_{\mu\nu}(x), \phi_{\mu\nu|\gamma}(x), \phi_{\mu\nu|\gamma\sigma}(x)) = 0] \Rightarrow [E_{\alpha\beta}(\bar{\phi}_{\mu\nu}(\bar{x}), \bar{\phi}_{\mu\nu|\gamma}(\bar{x}), \bar{\phi}_{\mu\nu|\gamma\sigma}(\bar{x})) = 0]$
However, $E_{\alpha\beta}(\phi_{\mu\nu}(x), \phi_{\mu\nu|\gamma}(x), \phi_{\mu\nu|\gamma\sigma}(x))$ vanishes identically for all values of the independent variables $x^1, \ldots, x^N$ so $E_{\alpha\beta}(\phi_{\mu\nu}(\bar{x}), \phi_{\mu\nu|\gamma}(\bar{x}), \phi_{\mu\nu|\gamma\sigma}(\bar{x})) = 0$ also holds. By the uniqueness of the solution we must therefore have $\phi_{\mu\nu}(\bar{x}) = \bar{\phi}_{\mu\nu}(\bar{x})$.

The boundary conditions have of course to be suitably chosen in order for the uniqueness condition to apply. Einstein considers a region Ω of space in which the energy tensor vanishes; the coordinates systems are such that they coincide outside and on the boundary of Ω but differ within Ω in a continuous way. We have now reached a contradiction, for, in general, the forms of the two functions $\phi_{\mu\nu}(\bar{x})$ and $\bar{\phi}_{\mu\nu}(\bar{x})$ are different; *i.e.* $\phi_{\mu\nu}(\bar{x}) \neq \bar{\phi}_{\mu\nu}(\bar{x})$.

Einstein expresses his and Grossmann's regret at having to give up covariance which they had set out to achieve. Grossmann's efforts however were later to prove very fruitful. He initiated the method of approaching the field equations via the Riemann–Christoffel and Ricci tensors, which are bound up with the geometrical properties of space. This is once again a situation in which mathematics played an essential heuristic role in the 'logic' of physical discovery. Once the physicist decides on the mathematical framework in which to formulate his theories, he can then let the framework guide him in the choice of the physically relevant quantities and relations. He thereby relies on the possibility that the mathematics has captured more of the physics than he had consciously intended. He lets the formalism do the work for him (see section 1 (*i*) above). In this particular case, after identifying the metric with the gravitational potential, Grossmann and Einstein rely on the mathematics (absolute differential calculus) to tell them which quantities are relevant from the geometrical and hence also from the physical point of view. (Curvature tensor, Ricci tensor, parallel transport of vectors).

Between 1913 and 1916 the progress in Einstein's thinking about Relativity can be viewed as a gradual and somewhat painful reclaiming of general covariance. Considering that he had no new 'facts' to guide him and had moreover reached a negative conclusion about the covariance principle, it was mainly faith in the correctness of Mach's position that sustained him: namely his belief in the view that no set of reference frames has a privileged status. The covariance principle he was effectively applying was far more stringent than the general one which, we have seen, turns out to be vacuous. He was looking for covariant differential equations involving only the $g_{\mu\nu}$'s and their first and second derivatives.[1]

Mach certainly exerted a psychological influence on Einstein, but it was the equivalence principle rather than Mach's philosophy which led Einstein to identify the metric with the field and hence harden the covariance requirement. It was also the use of Poisson's equation as a limiting case which made him search for a solution containing at most second order derivatives of the $g_{\mu\nu}$'s. We shall later see in what sense these two requirements uniquely determine the field equations. Hilbert pointed out that, in virtue of $G^{\mu\nu}_{\|\nu} = 0$,[2] the system of equations $R_{\mu\nu} = 0$ is underdetermined. If the field equations are initially satisfied, then the further development of the system is governed by six equations only; this allows for the required arbitrariness in the choice of the coordinate system. (Four redundant equations corresponding to the choice of four coordinates.) This situation is analogous to the case in elementary algebra where we have to solve two equations whose left-hand sides are not independent; *e.g.* $2x+7y = a$ and $6x+21y = b$. This system may be impossible (namely if $b \neq 3a$). However, if the system holds for one pair (x_0, y_0), then: $b = 6x_0+21y_0 = 3\ (2x_0+7y_0) = 3a$. Consequently one equation is redundant and there exists an infinite number solutions governed by $2x+7y = a$; x can be chosen arbitrarily and y takes on the value $(a-2x)/7$.

Thus Einstein's general argument against covariance lost its force. However, there remained the problem of whether the specific equations envisaged by Grossmann, namely $R_{\mu\nu} = -kT_{\mu\nu}$, are acceptable physical laws. In 1915 Einstein published four papers[3] which culminated with the field equations as we know them today (namely $R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R = -kT_{\mu\nu}$ or $R_{\mu\nu} = -k(T_{\mu\nu}-\frac{1}{2}g_{\mu\nu}T)$). A key role in this development was played by the mathematical techniques of Riemannian Geometry and by the Special-Relativistic equation '$E = mc^2$'.

The first paper published on 4 November contains a major revision of the basic concepts and definitions of Riemannian Geometry. The class of allowable frames of reference is restricted and the definition of covariant divergence altered. The paper is a little bewildering, for not until the end does the author give any clear reason for the drastic steps he takes. Instead of following Einstein's rather confusing order of presentation, I shall proceed to a reconstruction which nonetheless remains faithful to the original work.

---

[1] For this important point I am indebted to Professor J. Stachel.

[2] $G^{\mu\nu}$ is Einstein's tensor $R^{\mu\nu}-\frac{1}{2}g^{\mu\nu}R$.

[3] See Einstein [1915 *a–d*].

The main difficulty arises from the fact that in $R^{\mu\nu} = -kT^{\mu\nu}$ the divergence of $T^{\mu\nu}$ but not in general that of $R^{\mu\nu}$ vanishes. $R^{\mu\nu} = -kT^{\mu\nu}$ can also be written as:

$$R^{\mu}_{\nu} = -kT^{\mu}_{\nu} \quad \text{and} \quad R_{\mu\nu} = -kT_{\mu\nu}.$$

Contracting λ and ν we obtain: $R = -kT$. Hence: $R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = -k(T^{\mu\nu} - \frac{1}{2}g^{\mu\nu}T)$. Taking the divergence of the two sides, noting that the divergence of $(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R)$ vanishes identically and that $T^{\mu\nu}_{\|\nu} = 0$, we have: $(g^{\mu\nu}T)_{\|\nu} = 0$; *i.e.* $g^{\mu\nu}\frac{\partial T}{\partial x^{\nu}} = 0$. Multiplying by $g_{\sigma\mu}$:

$$\delta^{\nu}_{\sigma}\frac{\partial T}{\partial x^{\nu}} = 0; \quad i.e. \quad \frac{\partial T}{\partial x^{\sigma}} = 0.$$

The last question holds for all $\sigma = 0,1,2,3$. T must consequently be constant. Einstein knew that this was not generally the case. The equation $R_{\mu\nu} = -kT_{\mu\nu}$ had therefore to be modified, for a mere limitation on the number of allowable reference frames does not affect the above argument. In the paper Einstein offers a drastic solution which is covariant only with respect to a subgroup of the set of all transformations. $R_{\mu\nu}$ is written as:

$$R_{\mu\nu} = K_{\mu\nu} + H_{\mu\nu};$$

where: (2)

$$K_{\mu\nu} = \left\{ {\alpha \atop \mu\beta} \right\} \left\{ {\beta \atop \nu\alpha} \right\} - \frac{\partial}{\partial x^{\alpha}} \left\{ {\alpha \atop \mu\nu} \right\}$$

and:

$$H_{\mu\nu} = \frac{\partial^2}{\partial x^{\mu}\partial x^{\nu}}(\log\sqrt{-g}) - \left\{ {\alpha \atop \mu\nu} \right\} \frac{\partial \log\sqrt{-g}}{\partial x^{\alpha}}.$$

Consider the group of all transformations $\bar{x}^{\mu} = \bar{x}^{\mu}(\bar{x}^1, \ldots, x^N)$ such that

$$D = \frac{\partial(\bar{x}^1, \ldots, \bar{x}^N)}{\partial(x^1, \ldots, x^N)} = 1. \quad \text{Hence: } \bar{g} = \det(\bar{g}_{\mu\nu}) = \det\left(\frac{\partial x^{\alpha}}{\partial \bar{x}^{\mu}} \cdot \frac{\partial x^{\beta}}{\partial \bar{x}^{\nu}}\, g_{\alpha\beta}\right)$$

$$= D^{-2}\det(g_{\alpha\beta}) = \det(g_{\alpha\beta}) = g.$$

It follows that g, hence also $\sqrt{-g}$, are invariant. ∂ *log* $\sqrt{-g}/\partial x^{\nu}$ is now a covariant vector and:

$$\left(\frac{\partial \log\sqrt{-g}}{\partial x^{\nu}}\right)_{\|\mu} = \frac{\partial^2}{\partial x^{\mu}\,\partial x^{\nu}}(\log\sqrt{-g}) - \left\{ {\alpha \atop \mu\nu} \right\} \frac{\partial \log\sqrt{-g}}{\partial x^{\alpha}} = H_{\mu\nu}$$

is a second order tensor. Since, by (2): $K_{\mu\nu} = R_{\mu\nu} - H_{\mu\nu}$, $K_{\mu\nu}$ is also a tensor, so the relations:

$$K_{\mu\nu} = -kT_{\mu\nu}, \tag{3}$$

which Einstein chose as his field equations, are covariant with respect to the group. Also, because g is an invariant, the definition of the covariant divergence can be simplified as follows: Let $A^{\mu\nu}$ be an arbitrary contravariant tensor of the second order.

$$A^{\mu\nu}_{\|\sigma} = \frac{\partial A^{\mu\nu}}{\partial x^{\sigma}} + \begin{Bmatrix} \mu \\ \alpha\sigma \end{Bmatrix} A^{\alpha\nu} + \begin{Bmatrix} \nu \\ \alpha\sigma \end{Bmatrix} A^{\mu\alpha}.$$

Therefore:

$$\begin{aligned} A^{\mu\nu}_{\|\nu} &= \frac{\partial A^{\mu\nu}}{\partial x^{\nu}} + \begin{Bmatrix} \mu \\ \alpha\nu \end{Bmatrix} A^{\alpha\nu} + \begin{Bmatrix} \nu \\ \alpha\nu \end{Bmatrix} A^{\mu\alpha} \\ &= \frac{\partial A^{\mu\nu}}{\partial x^{\nu}} + \begin{Bmatrix} \mu \\ \alpha\nu \end{Bmatrix} A^{\alpha\nu} + \frac{\partial \log \sqrt{-g}}{\partial x^{\alpha}} A^{\mu\alpha}; \end{aligned}$$

hence:

$$A^{\mu\nu}_{\|\nu} - \frac{\partial \log \sqrt{-g}}{\partial x^{\alpha}} A^{\mu\alpha} = \frac{\partial A^{\mu\nu}}{\partial x^{\nu}} + \begin{Bmatrix} \mu \\ \alpha\nu \end{Bmatrix} A^{\alpha\nu}$$

Again, since $\frac{\partial \log \sqrt{-g}}{\partial x^{\mu}}$ is a covariant and hence $\frac{\partial \log \sqrt{-g}}{\partial x^{\mu}} A^{\mu\alpha}$ is a contravariant vector, both sides of the last equation are contravariant vectors. We can therefore define a new covariant divergence $A^{\mu\nu}:\nu$ as follows:

$$A^{\mu\nu}:\nu \underset{\text{def.}}{=} \frac{\partial A^{\mu\nu}}{\partial x^{\nu}} + \begin{Bmatrix} \mu \\ \alpha\nu \end{Bmatrix} A^{\alpha\nu}$$

An important point to note is that the set of reference frames must be such that g assumes a constant value in none of them; otherwise $\frac{\partial \log \sqrt{-g}}{\partial x^{\alpha}}$ vanishes, and so does $H_{\mu\nu}$, by (2); therefore the two equations

$$R_{\mu\nu} \equiv K_{\mu\nu} + H_{\mu\nu} = -k\,T_{\mu\nu} \quad \text{and} \quad K_{\mu\nu} = -k\,T_{\mu\nu}$$

coincide, and Einstein had already found the first equation unacceptable.

Einstein was obviously dissatisfied with the first desperate solution, in which the geometrical significance of the matrix $K_{\mu\nu}$ remains unclear. An alternative solution, which appeared on 11 November 1915, consisted in trying to get round the difficulty presented by the non-constant character of $T = T^{\nu}_{\nu}$. One important case in which the trace T vanishes identically is that of the electromagnetic field where:

$$T^{\mu}_{\nu} = -F^{\mu\beta}F_{\nu\beta} + \tfrac{1}{4}\,\delta^{\mu}_{\nu}\,F^{\alpha\beta}F_{\alpha\beta}.$$

(In the case of incoherent matter on which no external forces act, we have:

$$T^{\mu\nu} = \rho_0 \frac{dx^\mu}{ds} \frac{dx^\nu}{ds}.$$

and $$T = g_{\mu\nu} T^{\mu\nu} = \rho_0$$

this is why T in general differs from 0.)

At this stage Einstein pinned his hopes on a complete reduction of physics to some—possibly extended—electromagnetic theory, in which the trace T of T^μ_ν would still vanish. Matter would of course have to be explained in purely electromagnetic terms. Thus this second solution of the 11 November consisted in going back to general covariance, and accepting the equation $R_{\mu\nu} = -kT_{\mu\nu}$, while requiring that T should vanish. Einstein's position at this stage bears a close resemblance to that of Lorentz, who tried to reduce mass to electromagnetic inertia, thus depriving the electron of all material mass.[1] This corresponds to putting $\rho_0 = 0$ in the energy tensor of incoherent matter. In a previous paper[2] I have argued that one advantage of Einstein over Lorentz was the former's agnosticism about the ultimate nature of inertia and of molecular forces. Einstein must have realised how risky it was to tie General Relativity to any specific theory about the nature of matter; for on the 25 November 1915 he gave the paper in which the breakthrough was at last achieved. In fact the article of the 4 November already contained the kernel of the final solution; and the first hints go back to Einstein's [1914] paper. There Einstein considered the Special-Relativistic conservation law $\partial T^{\mu\nu}/\partial x^\nu \equiv T^{\mu\nu}_{|\nu} = 0$ which, on going over to curved space, has to be replaced by $T^{\mu\nu}_{\|\nu} = 0$ (or equivalently $T^{\nu}_{\mu\,\|\nu} = 0$). He remarks that this last equation is no longer a strict conservation law; in it the gravitational energy does not explicitly appear because it remains imbedded in the geometry. Hence arises the problem of determining a gravitational energy matrix t^ν_μ such that T^ν_μ and t^ν_μ *together* are conserved; *i.e.* such that $\partial(T^\nu_n + t^\nu_a)/\partial x^\nu = 0$.

In his [1915a] Einstein used variational methods in order to derive the field equations and thereby identified the quantity t^ν_μ as:

$$(4) \quad t^\nu_\mu = \tfrac{1}{2}\,\delta^\nu_\mu g^{\lambda\sigma}\,\Gamma^\alpha_{\lambda\beta}\,\Gamma^\beta_{\sigma\alpha} - g^{\lambda\sigma}\,\Gamma^\alpha_{\lambda\mu}\,\Gamma^\nu_{\sigma\alpha} \qquad \left(\Gamma^\alpha_{\beta\gamma} = -\left\{{\alpha \atop \beta\gamma}\right\}\right)$$

The field equations could then be written as:

$$(5) \quad \frac{\partial}{\partial x^\alpha}\left(g^{\sigma\nu}\,\Gamma^\alpha_{\mu\sigma}\right) - \tfrac{1}{2}\,\delta^\nu_\mu\, g^{\gamma\sigma}\,\Gamma^\alpha_{\gamma\beta}\,\Gamma^\beta_{\sigma\alpha} = -k\,T^\nu_\mu - t^\nu_\mu$$

Let us now turn to Einstein's [1915*d*]. Again all frames of reference are to be allowed, so we can choose a coordinate system in which $g = -1$ at all points. Einstein takes $R_{\mu\nu} = 0$ to be the correct law for empty space, because these equations had enabled him earlier that year to explain the precession of Mercury's perihelion.[1] Since he no longer wanted to assume $T = 0$, he had to modify the relation $R_{\mu\nu} = -kT_{\mu\nu}$ but had first to determine where the fault lay. Let us note that, since g is constant, $\frac{\partial \log \sqrt{-g}}{\partial x^{\alpha}} = 0$; hence, $H_{\mu\nu} = 0$ and $R_{\mu\nu} = K_{\mu\nu}$ (by equation 2)). Thus the field equation $R_{\mu\nu} = -kT_{\mu\nu}$ reduces to $K_{\mu\nu} = -kT_{\mu\nu}$. Because of this judicious choice of the coordinate system, from the strictly technical viewpoint we have to deal with the same mathematical relations as in Einstein's [1915*a*]; so the results achieved there can be brought to bear on the present problem.

---

[1] See Zahar [1973 *a,b*].

[2] *Ibid.*

With some rational reconstruction, Einstein's reasoning seems to have been as follows. The transition from $R_{\mu\nu} = 0$ to $R_{\mu\nu} = -kT_{\mu\nu}$ was made on the assumption that we go from zero to positive energy levels. This is mistaken, in view of the law identifying mass with energy ($E = mc^2$). $R_{\mu\nu} = 0$ applies to space empty of all *but* gravitational energy. However, since gravitational energy is present, it contributes to inertial mass, hence, by the equivalence principle, to gravitational mass as well. We can roughly say that the field is self-generating in the sense that its energy acts on one of its sources. (This explains the non-linear character of the field equations.) Einstein already had a gravitational energy matrix t^{ν}_{μ}. t^{ν}_{μ} was obtained by means of a standard mathematical method which was classically known to yield conservation laws, namely the variational method. Einstein considered a frame throughout which $\sqrt{-g} = 1$, *i.e.* $g = -1$. He then required: $\delta(\int H \, . \, dw) = 0$, where $dw = \sqrt{-g} \, . \, dx^0 \, . \, dx^1 \, . \, dx^2 \, . \, dx^3 = dx^0 \, . \, dx^1 \, . \, dx^2 \, . \, dx^3$, and $H = g^{im} \, . \, \Gamma^{\nu}_{i\mu} \, . \, \Gamma^{\mu}_{m\nu}$. In this way he obtained both the field equations for empty space,

$$\frac{\partial \Gamma^{\sigma}_{\mu\nu}}{\partial x^{\sigma}} + \Gamma^{j}_{\mu l} \, . \, \Gamma^{l}_{\nu j} \equiv \frac{\partial}{\partial x^{\sigma}} \left(\frac{\partial H}{\partial g^{\mu\nu}_{\sigma}} \right) - \frac{\partial H}{\partial g^{\mu\nu}} = 0$$

and the conservation law $t^{\nu}_{\mu|\nu} = 0$.

Because of this conservation law, *i.e.* because of a mathematical equation which usually expresses the constancy through time of a substance, he interpreted the matrix t^{ν}_{μ} as referring to gravitational energy; this, despite the fact that t^{ν}_{μ} is frame-dependent, hence should normally represent nothing physically significant in Relativity Theory.

Let me now describe how Einstein arrived at field equations for the case where non-gravitational forms of energy are present. One way of obtaining new equations is to rewrite $R_{\mu\nu} = 0$ in a way that exhibits the dependence of the field on its energy, *i.e.* on t^{ν}_{μ}. We can then generalise the equation by adding kT^{ν}_{μ} to t^{ν}_{μ} (and kT to t); kT^{ν}_{μ} and t^{ν}_{μ} have to play analogous roles, which is certainly not the case if we adopt $R_{\mu\nu} = -kT_{\mu\nu}$ as our field equations. This can be seen by contracting ν and μ in (4) and (5). We obtain:

---

[1] See Einstein [1915*c*].

$$t = t^{\mu}_{\mu} = \tfrac{1}{2}\, g^{\lambda\sigma} \,.\, \Gamma^{\alpha}_{\lambda\beta} \,.\, \Gamma^{\beta}_{\sigma\alpha} \qquad {}^{\alpha}_{\lambda\mu} \,.\, \Gamma^{\mu}_{\sigma\alpha} = g^{\lambda\sigma} \,.\, \Gamma^{\alpha}_{\lambda\beta}\, \Gamma^{\beta}_{\sigma\alpha}$$

Substituting in (5):

$$(5') \quad \frac{\partial}{\partial x^{\alpha}}(g^{\sigma\nu}\, \Gamma^{\alpha}_{\mu\sigma}) - \tfrac{1}{2}\, \delta^{\nu}_{\mu}\, t = -kT^{\nu}_{\mu} - t^{\nu}_{\mu}$$

It is obvious that, in (5′), kT^{ν}_{μ} and t^{ν}_{μ} play totally assymetrical roles. Dissatisfaction with this anomaly started Einstein on his search for different field equations.

In fact the solution is near at hand if we remember that (5′) is supposed to hold for empty space, *i.e.* for the case where $T^{\nu}_{\mu} = 0$. Hence:

$$(6) \quad \frac{\partial}{\partial x^{\alpha}}(g^{\sigma\nu} \,.\, \Gamma^{\alpha}_{\mu\sigma}) - \tfrac{1}{2}\delta^{\nu}_{\mu} \,.\, t = -t^{\nu}_{\mu};$$

$$i.e. \quad \frac{\partial}{\partial x^{\alpha}}(g^{\sigma\nu} \,.\, \Gamma^{\alpha}_{\mu\sigma}) = -t^{\nu}_{\mu} + \tfrac{1}{2}\, \delta^{\nu}_{\mu}\, t$$

In this equation, which is equivalent to $R_{\mu\nu} = 0$, the dependence of the field on the gravitational energy is clearly exhibited. For the case where extra energy represented by T^{ν}_{μ} is present, we generalise (6) by adding kT^{ν}_{μ} to t^{ν}_{μ} and kT to kt, where k is some constant. Thus we obtain:

$$\frac{\partial}{\partial x^{\alpha}}(g^{\sigma\nu}\, \Gamma^{\alpha}_{\mu\sigma}) = -(t^{\nu}_{\mu} + k \,.\, T^{\nu}_{\mu}) + \tfrac{1}{2}\delta^{\nu}_{\mu}\, (t + kT).$$

Working back from these equations we arrive at:

$$(7) \quad \begin{cases} R^{\nu}_{\mu} = -k\, (T^{\nu}_{\mu} - \tfrac{1}{2}\, \delta^{\nu}_{\mu} T); \quad \text{or} \quad R_{\mu\nu} = -k\, (T_{\mu\nu} - \tfrac{1}{2}\, g_{\mu\nu}\, T); \quad \text{or} \\ R_{\mu\nu} - \tfrac{1}{2}\, g_{\mu\nu}\, R = -kT_{\mu\nu}. \end{cases}$$

These are the field equations as they are generally accepted today.

Two philosophical points can be made. First, this process of discovery illustrates once again the crucial role played by '$E = mc^2$' in the genesis of *GTR*, thus underlining the continuity which exists between the Special and the General Theories. The second point is that the heuristic function of mathematics in physical discovery can hardly be exaggerated in this case: the most important step consisted in isolating a gravitational energy matrix t^{ν}_{μ}. Although 'energy' is a physical concept, the fact that t^{ν}_{μ} is not a tensor, and hence has no invariance properties, makes its physical interpretation highly problematic; its main right to being called an *energy* pseudo-tensor stems from the vanishing of the (ordinary) divergence $\partial(T^\nu_\mu + t^\nu_\mu)/\partial x^\nu$, *i.e.* from a strictly mathematical property. Once t^ν_μ had been singled out, there remained the mathematical problem of rewriting $R_{\mu\nu} = 0$ so as to bring out the energy term t^ν_μ, then to generalise the equation by adding kT^ν_μ to t^ν_μ (and kT to t). However, an important qualification has to be made about the role of mathematics in scientific discovery. Einstein could after all have proceeded in the way which many textbooks use for purposes of exposition: realising that the divergence of $R^{\mu\nu}$ does not identically vanish, he might have tried to add extra terms to the left-hand side of $R^{\mu\nu} = -kT^{\mu\nu}$ so as to obtain a divergenceless tensor. In fact such a method would have been mathematically sound, since $(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \lambda g_{\mu\nu})$ is the most general divergenceless tensor satisfying the condition of being a function only of the $g_{\mu\nu}$'s and of their first and second derivatives.[1] That Einstein did not resort to such manipulative tricks confirms the view that pure mathematical gimmickry played no significant role in the development of the Relativity programme.[2] Einstein was concerned about the physics and not about the mathematical machinery. However, having said this, I still maintain that one can distinguish those situations in which mathematics rather than physical intuition played a dominant heuristic role· An example will illustrate this distinction. In classical mechanics Descartes had an intuitive notion of the momentum as a quantity proportional both to velocity and to quantity of matter. Newton, using the same concept of momentum, laid down three postulates from which the conservation law of momentum was deduced. The mathematical formalism captured concepts, properties and relations which could be intuitively grasped; the conservation laws then follow from the axioms connecting the basic concepts and relations. In Einstein's case at this stage of the programme, exactly the reverse process occurred; he is already in possession of a mathematical system; the system guides him in isolating relevant physical quantities such as t^ν_μ, which are mathematically—not intuitively—characterised by a conservation law. This example illustrates the second heuristic role which mathematics can play in physical discovery (see above section 1 (*ii*)). Through physically interpreting a hitherto uninterpreted—or rather a seemingly uninterpretable—mathematical entity like t^ν_μ, Einstein was led to a physical discovery. Einstein's position is analogous to Lorentz's, who was led by the mathematical structure of Maxwell's equations to the Lorentz transformation and then to the contraction hypothesis.[3]

---

[1] Cartan first established this result for the case where the tensor is linear in the second derivatives. Professor J. Stachel pointed out to me that the linearity can be omitted.

[2] See Zahar [1973*b*].

[3] See Zahar [1973*b*].

(*v*) *The Negative Heuristic*

We have seen that the covariance principle, if looked upon as an ordinary proposition, is either vague and metaphysical or else precise and vacuous[1]; only through being considered as part of the positive heuristic can its regulative role be appreciated. Does this entail that the Relativity programme has no hard core? One law which runs through all the later stages of the programme qualifies as its negative heuristic, namely the law about the interchangeability of mass and energy ($E = mc^2$). We have seen how this law forced Einstein first to abandon Special Relativity and then much later enabled him to alter his equation $R_{\mu\nu} = -kT_{\mu\nu}$ and obtain $R_{\mu\nu} = -k(T_{\mu\nu}-\frac{1}{2}g_{\mu\nu}T)$. Apart from its technical importance this law has a deep metaphysical significance. Traditionally matter had been considered the underlying substance of the world and energy a mode of existence of matter. '$E = mc^2$' shows that matter and energy are continuous with each other, or rather that they constitute one and the same thing. With this new law we have a complete revision of our way of looking at the physical world. The reason why I hesitate a little about taking this law to be the hard core of the programme is that only with hindsight has its importance been fully appreciated. Unlike the covariance principle, it was not a central tenet of Einstein's in his [1905]. Later on in 1905 Einstein realised that the inertia of a body may depend on its energy content; again in 1907 he gave two different 'proofs' of $E = mc^2$. It seems therefore that the full significance of this law was gradually realised. However, if we allow the negative heuristic gradually to emerge as the programme unfolds—and Lakatos did allow for this possibility—we can look upon $E = mc^2$ as the hard core of relativity; or rather, instead of this precise equation, we take the more qualitative statement that mass and energy are interchangeable.

The following diagram indicates the way in which the development of Relativity Theory can be accounted for within the Methodology of Scientific Research Programmes.

NEGATIVE HEURISTIC (Hard Core)

1. Relativity Principle (Covariance Principle)
2. Interchangeability of Mass and Energy ($E = mc^2$)

POSITIVE HEURISTIC

(1) Lorentz-covariance. (Einstein applies the Lorentz-covariance principle to Maxwell's equations;	(1′) General Covariance together with the unity requirement (Einstein and the 'minimal'

---

[1] See, however, Appendix.

Planck applies it to Newton's second law; Einstein to the law of conservation of energy; Lewis and Tolman to the laws of conservation of momentum and energy.)	coupling principle; Weyl's and Eddington's attempts at unified field theories).
(2) Use of old laws as limiting cases of the new relativistic laws. (Use of Newton's second law; of the conservation laws; of Maxwell's equations. (Maxwell's equations are retained).)	(2′) Use of old laws as limiting cases (Poisson's equation and Special Relativity).
(3) Analysis of the well-known 'fact' that electromagnetic induction depends on relative and not on absolute motion.	(3′) Analysis of the well-known 'fact' that all bodies fall with the same acceleration.
	(4′) Heuristic function of the Principle of Equivalence. (An inertial field is nothing but a special (reducible) case of a gravitational field.)

APPENDIX: THE STATUS OF THE RELATIVITY PRINCIPLE

The problem concerning the Relativity—or Covariance—principle is as follows. Let $\mathscr{C}$ be any statement. The Relativity Principle $R[\mathscr{C}]$ is the proposition that, if $\mathscr{C}$ is any law of nature, then $\mathscr{C}$ is covariant. The main objection (first raised by Kretschmann) to the Relativity Principle is that any physical theory can be written in generally covariant form. In other words: given any physical law $\mathscr{C}$, there always exists a generally covariant physical law $\mathscr{C}^*$ such that $\mathscr{C}\Leftrightarrow\mathscr{C}^*$. Thus: $R[\mathscr{C}^*]$, since $\mathscr{C}^*$ is generally covariant.

If we consider $R[\mathscr{C}]$ as an object-language statement in which $\mathscr{C}$ occurs, then from $\mathscr{C}\Leftrightarrow\mathscr{C}^*$ and $R[\mathscr{C}^*]$, there follows $R[\mathscr{C}]$. Since $\mathscr{C}$ is an arbitrary physical law, $R[\mathscr{C}]$ is vacuous. Even if we accept (Kretschmann's) thesis that, for any $\mathscr{C}$, there exists a generally covariant $\mathscr{C}^*$ such that $\mathscr{C}\Leftrightarrow\mathscr{C}^*$, the above argument goes through only if $R[\mathscr{C}]$ is treated as a statement in the object-language. But we know that the Relativity Principle is not vacuous, since we can exhibit statements (*e.g.* $\sum m_0\vec{v} = 0$) which are not Lorentz-covariant, let alone generally covariant. Let me therefore propose the following—very tentative—solution to the problem posed by the status of the Relativity Principle.

The very formulation of the Relativity principle indicates that it is a proposition about the *statement* ‘$\mathscr{C}$’; in other words, the Relativity principle belongs to the metalanguage and should be written as $R(\text{‘}\mathscr{C}\text{’})$. Instead of being an object-language statement $R[\mathscr{C}]$ in which $\mathscr{C}$ *occurs*, the Relativity Principle is a metalinguistic statement *about* the statement ‘$\mathscr{C}$’. Thus, although $\mathscr{C}\Leftrightarrow\mathscr{C}^*$ and $R(\text{‘}\mathscr{C}^*\text{’})$, we cannot conclude that $R(\text{‘}\mathscr{C}\text{’})$, *i.e.*, since ‘$\mathscr{C}$’ $\neq$ ‘$\mathscr{C}^*$’, we cannot infer $R(\text{‘}\mathscr{C}\text{’})\Leftrightarrow R(\text{‘}\mathscr{C}^*\text{’})$. Putting it a little differently—if $R[\mathscr{C}]$ *were* a statement in the object-language, then, from $\mathscr{C}\Leftrightarrow\mathscr{C}^*$ there follows $R[\mathscr{C}]\Leftrightarrow R[\mathscr{C}^*]$, hence, since $R[\mathscr{C}^*]$, we must also have $R[\mathscr{C}]$. $R[\mathscr{C}]$ however is *not* part of the object-language and should be written $R(\text{‘}\mathscr{C}\text{’})$. The upshot of all this is that, although ‘$\mathscr{C}$’ can be written in the generally covariant form ‘$\mathscr{C}^*$’, ‘$\mathscr{C}$’ *itself* is not generally covariant.

Remember that $R(\text{‘}\mathscr{C}\text{’})$ stands for: if ‘$\mathscr{C}$’ is a law of nature, then ‘$\mathscr{C}$’ is generally covariant. Let $\mathscr{R}$ stand for: $R(\text{‘}\mathscr{C}\text{’})$ for all ‘$\mathscr{C}$’. Let $\mathscr{R}'$ stand for: for all ‘$\mathscr{C}$’, if ‘$\mathscr{C}$’ is a law of nature, then ‘$\mathscr{C}$’ is Lorentz-covariant. Obviously: $\mathscr{R}\Rightarrow\mathscr{R}'$. $\mathscr{R}'$ can be considered as part of the hard core of the Relativity Programme. General Relativity strengthened $\mathscr{R}'$ into $\mathscr{R}$.

London School of Economics

REFERENCES

Becker, O. [1963]: *Grundlagen der Mathematik.*

Bridgman, P. W. [1936]: *The Nature of Physical Theory.*

Duhem, P. [1906]: *La Théorie Physique: Son Objet, Sa Structure.*

Eddington, A. S. [1920]: *Space, Time and Gravitation.*

Eddington, A. S. [1924]: *The Mathematical Theory of Relativity.*

Einstein, A. [1905]: ‘Zur Elektrodynamik bewegter Körper’, *Annalen der Physik,* **17**, pp. 891–921.

Einstein, A. [1912]: ‘Relativität und Gravitation. Erwiderung auf eine Bemerkung von M. Abraham’, *Annalen der Physik,* **38**, 1059–64.

Einstein, A. [1914]: ‘Die formale Grundlage der allgemeinen Relativitätstheorie’, *Sitzungsberichte der preussischen Akademie der Wissenschaften,* 1914, part 2, pp. 1030–85.

Einstein, A. [1915*a*]: ‘Zur allgemeinen Relativitätstheorie’, *Sitzungsberichte der preussischen Akademie der Wissenschaften,* 1915, part 2, pp. 778–86.

Einstein, A. [1915*b*]: ‘Zur allgemeinen Relitivitäts theorie (Nachtrag)’, *Sitzungsberichte der preussischen Akademie der Wissenschaften,* 1915, part 2, pp. 799–801.

Einstein, A. [1915*c*]: ‘Erklärung der Perihelbewegung des Merkurs aus der allgemeinen Relativistätstheorie’, *Sitzungsberichte der preussischen Akademie der Wissenschaften,* 1915, part 2, pp. 831–9.

Einstein, A. [1915*d*]: ‘Die Feldgleichungen der Gravitation’, *Sitzungsberichte der preussischen Akademie der Wissenschaften,* 1915, part 2, pp. 844–7.

Einstein, A. [1915*e*]: Letter to Sommerfeld 28 Nov. 1915; in A. Hermann (*ed.*): *Albert Einstein und Arnold Sommerfeld: Briefwechsel* 1968, pp. 32–6.

Einstein, A. [1916]: ‘Grundlage der Allgemeinen Relativistätstheorie’, *Annalen der Physik,* **49**, 768–822. Page reference is to the English translation ‘The Foundation of the General Theory of Relativity’ in A. Einstein and others: *The Principle of Relativity,* pp. 116–64.

EINSTEIN, A. [1918]: 'Prinzipielles zur Allgemeinen Relativitätstheorie', *Annalen der Physik*, **55**, 241–4.

EINSTEIN, A. [1922]: *The Meaning of Relativity.*

EINSTEIN, A. and others [1923]: *The Principle of Relativity.*

EINSTEIN, A. [1934]: *Mein Weltbild.*

EINSTEIN, A. [1950]: *Out of My Later Years.*

EINSTEIN, A. and GROSSMANN, M. [1913]: 'Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation', *Zeitschrift für Mathematik und Physik*, **62**, 225–61.

FEYNMAN, R. [1965]: *The Character of Physical Law.*

HILBERT, D. and COURANT, R. [1924]: *Methoden der mathematischen Physik.*

KRETSCHMANN, E. [1917]: 'Über den physikalischen Sinn der Relativitätspostulate, A. Einstein's neue und seine ursprüngliche Relativitätstheorie', *Annalen der Physik*, **53**, 575–614.

LEWIS, G. N. and TOLMAN, R. C. [1909]: 'The Principle of Relativity and non-Newtonian Mechanics', *Philosophical Magazine*, Series 5, 18, pp. 510–23.

MACH, E. [1883]: *Die Mechanik in ihrer Entwicklung, historisch-kritisch dargestellt.*

MACH, E. [1905]: *Erkenntnis und Irrtum.*

MACH, E. [1909]: *The History and the Root of the Principle of the Conservation of Energy.*

MEYERSON, E. [1908]: *Identité et Réalitié.*

MEYERSON, E.]1921]: *De L'Explication dans les Sciences.*

MEYERSON, E. [1924]: *La Déduction Relativiste.*

POPPER, K. R. [1963]: *Conjectures and Refutations.*

POPPER, K. R. [1972]: *Objective Knowledge.*

REDHEAD, M. L. G. [1975]: 'Symmetry in Intertheory Relations', *Synthese*, **32**.

RIEMANN, B. [1953]: *Collected Works.*

SCHRÖDINGER, E. [1950]: *Space-Time Structure.*

STACHEL, J. [1979]: 'Einstein and the Rigidly Rotating Disc' (forthcoming).

TRAUTMANN, A., PIRANI, F. A. E. and BONDI, H. [1964]: *Lectures on General Relativity.*

WEYL, H. [1923]: *Raum, Zeit, Materie.*

WEYL, H. [1949]: *Philosophy of Mathematics and Natural Science.*

ZAHAR, E. G. [1973*a*]: *The Development of Relativity Theory: A Case Study in the Methodology of Scientific Research Programmes.* (Doctoral Dissertation, University of London.)

ZAHAR, E. G. [1973*b*]: 'Why did Einstein's Programme supersede Lorentz's?' *The British Journal for the Philosophy of Science*, **24**, pp. 95–123 and pp. 223–62.

ZAHAR, E. G. [1977]: 'Mach, Einstein and the Rise of Modern Science', *The British Journal for the Philosophy of Science*, **28**, pp. 195–213.

Brit. J. Phil. Sci. 31 (1980), 45–62 *Printed in Great Britain* 45

Bayesianism versus Baconianism in the Evaluation of Medical Diagnoses

by L. JONATHAN COHEN

1 Medicine is an excellent example of a field in which it is often necessary to make inferences under conditions of uncertainty, either because further delays in therapy would be too risky or because further tests or examinations would be too costly. In the words of Sardeman [1964], p. 142, 'If you did all laboratory tests to everybody you would never get finished, and nobody would be able to pay for it.' But some uncertain inferences are less reliable than others, and it is important to be able to draw the right distinctions here. So how is diagnostic uncertainty to be graded? Two models for the evaluation of medical diagnoses are often found in apparent conflict with one another today. The first is avowedly Bayesian. The second is demonstrably Baconian in its logic though not normally conceiving of itself as such. The purpose of the present paper is to clarify the nature of the difference between these two models and to show that, when restricted to their legitimate roles, they do not conflict with one another.

Some initial clarification of terminology may be helpful. By 'evidence' I shall mean one or more facts of the following four kinds: (1) facts about the patient's social or physical environment, or his occupation, or life-style, or biography, (2) symptoms, *i.e.* the patient's complaints, (3) signs, *i.e.* objectively observable manifestations of the patient's state of health that the physician himself can detect, and (4) the results of laboratory tests, of roentgenography or of any other process that generates data from the patient's body with the aid of paraclinical procedures. (All four kinds of evidential facts may be heavily 'theory-laden', in the sense that their present recognition in particular cases is the result of previous scientific effort in regard to classes of similar cases.) By a 'disease' I shall mean any causal process within a person's body whereby one or more symptoms and/or signs of ill health is produced. Of course, many such causal processes are initiated from without—by a blow on the head, say, or by the invasion of an infecting parasite, or by contact with a carcinogenic substance, *etc.* But I am confining my attention here to diagnoses that are instrumental for therapy, as distinct from those appropriate to preventive medicine. So it is convenient to restrict the concept of diseases

Received 3 *April* 1979

here to causal processes within the body. Again, there are some commonly associated groups of symptoms and/or signs of ill health for which no satisfactory aetiology is yet available. I shall refer to them as 'syndromes', and assume that any report of them in a particular patient is a description, not a diagnosis.

In these terms the progress of pathology consists in replacing the description of syndromes by the diagnosis of diseases. But this process of replacement is an open-ended one, as the chain of causation is traced backwards. Modern nosology begins in the seventeenth century with Sydenham's conception of syndromes as natural species; it makes a great leap forward in the nineteenth century when the practice of autopsy begins to reveal macro-anatomic causes of syndromes and thus confirms the scientific respectability of conceiving a disease as a causal process; it makes another great leap forward in the following decades when the vast proliferation of paraclinical tests makes it possible to trace many disease-processes further back and permits the classification of many diseases in terms of microentities or microstructures that are studied within newly specialised disciplines (bacteriology, virology, haematology, immunology, *etc.*); and one can see that yet another great leap forward is conceivable if many more diseases (beyond the familiar industrial and nutritional ones) could be traced back to the social or environmental factors that foster or provoke them. A nosology is thus conceivable in which, for the purposes of preventive medicine, it would be convenient to conceive of all disease as a causal process that includes social and environmental events as well as bodily ones, and it would then be right to modify the classification of diseases accordingly. But in the present state of knowledge, and for the purposes of most curative medicine, a more restricted conception of disease is appropriate. Even then the causal end-point of diagnostic analysis and nosological definition is often not determined uniquely. To cite an example given by Feinstein [1973], p. 217, is myocardial infarction a satisfactory end-point for an appropriate type of chest-pain, or should the series of admissible diagnoses extend onward from myocardial infarction to coronary thrombosis, coronary atherosclerosis, hyperlipemia, nutritional imbalance and neurotic personality?

2 The Bayesian theory appears in various forms. What is essential to it is that the validity of a diagnosis is thought of as a conditional probability—the probability that the patient has a certain specified disease, given certain specified evidence—and that this conditional probability is taken to be a function partly of the probability of the evidence, given that he has the disease, and partly of relevant prior probabilities. In its simplest and most

straightforward form the theory is just exploiting the familiar principle in the mathematical calculus of chance

$$p(D/E) = \frac{p(E/D) \times p(D)}{p(E)}$$

where $p(E) > 0$, with 'D' standing for the name of the diagnosed disease and 'E' for the record of the evidence.

But as some Bayesian writers on medical diagnosis, *e.g.* Ledley and Lusted [1959], point out it is possible to eliminate $p(E)$ here, so long as it is safe to assume that in the particular case under examination E is caused by just one of the diseases D_1, D_2, . . ., D_n. If D_k is now the diagnosed disease we can write

$$p(D_k/E) = \frac{p(E/D_k) \times p(D_k)}{\sum_{i=1}^{n} p(E/D_i) \times p(D_i)}$$

An arguable merit of the theory is thus that it directs a clinician's attention to the importance of taking into account the value of $p(D)$ that is appropriate for his own patients: some kinds of disease may be much commoner among those already admitted to a hospital, for instance, than among those who consult their general practitioner, as Wulff [1976], pp. 83–4 points out.

A number of objections have been made to the proposal of Bayesian evaluations for medical diagnoses, *e.g.* Feinstein [1973], 276 ff., Lange [1976], p. 16, Jacquez and Norusis [1976], p. 381, and Feinstein [1977]. These objections fall into four main categories.

The first type of objection argues that a Bayesian account seeks merely to determine the probability with which a certain nosological term fits a given array of assembled data. Bayesians are therefore accused of being blind both to the importance of understanding the causal processes that have generated the data and also to the sequential nature of diagnostic procedures and their frequent incorporation of feedback, whereby initial examinations suggest tests, test-results suggest therapy, a patient's positive or negative response to therapy may suggest further lines of investigation, and so on. The situation to be represented is very much more complex than the Bayesian formula is capable of representing, and in any case clinical reasoning often terminates in therapeutic action rather than diagnostic conclusion.

But the Bayesian has a ready answer to at least part of this objection. He merely points out that 'D' in '$p(D/E)$' can stand for any kind of event, property or process that may be inferred, at the appropriate level of probability, from E, where E is the evidence available at any one stage of the diagnostic investigation. So a sequential form of Bayes' theorem can represent the sequential nature of diagnostic procedures. Obviously a Bayesian account could have no right to claim that it revealed heuristic strategies which are specific to the medical situation, such as the investigation of underlying causal processes or the interspersion of diagnosis with therapy. But that would not debar it from claiming an ability to grade the probability of the inference in which any such strategy issues. Moreover, a Bayesian account here is on much safer ground if offered as an idealised reconstruction of the relevant clinical evaluations rather than as a practically applicable methodology for making them, though in this relatively unambitious form it may be of greater interest to philosophers than to physicians.

A second type of objection stresses the risks incurred in assuming that in the particular case under examination the pathological manifestations result from just one of the diseases $D_1, D_2, \ldots, D_n$. Suppose yet another disease could cause such manifestations, or suppose that in the particular case they were caused by the presence of two or more of $D_1, D_2, \ldots, D_n$?

But the Bayesian might legitimately reply that we can only take into account what we already know. If additional disease-candidates become available we shall naturally have to revise our estimates. Similarly if two or more of $D_1, D_2, \ldots, D_n$ can co-occur, the expression that may replace $p(E)$ in Bayes' law becomes correspondingly more complicated. We have to consider not $\sum_{i=1}^{n} p(E/D_i) \times p(D_i)$ but the sum of

$$p(D_1 \,\&\, \overline{D_2 \vee D_3 \vee \ldots \vee D_n}/E) \times p(D_1 \,\&\, \overline{D_2 \,\&\, D_3 \vee \ldots \vee D_n})$$

and

$$p(D_1 \,\&\, D_2 \,\&\, \overline{D_3 \vee D_4 \vee \ldots \vee D_n}/E) \times p(D_1 \,\&\, D_2 \,\&\, \overline{D_3 \vee D_4 \vee \ldots \vee D_n})$$

and so on.

A third type of objection argues that a Bayesian account is only able to accommodate the corroborative force of further evidence by assuming that each evidential fact is independent of every other, whereas such independence is rarely, if ever, to be found in diagnostic data. This objection is not normally worked out in detail. But if expressed in exact terms it would presumably run somewhat as follows. Or at any rate the objection seems to lack plausibility unless this is what it means.

An instance of corroboration may be represented by the inequality $p(D/E_1 \,\&\, E_2) > p(D/E_1)$. Here E_2 corroborates E_1's support for D. In Bayesian terms we therefore need to seek the conditions under which

$$\frac{p(E_1 \,\&\, E_2/D) \times p(D)}{p(E_1 \,\&\, E_2)} > \frac{p(E_1/D) \times p(D)}{p(E_1)}$$

These conditions will obviously be the same as those under which

$$(1) \qquad \frac{p(E_1 \,\&\, E_2/D)}{p(E_1 \,\&\, E_2)} > \frac{p(E_1/D)}{p(E_1)}$$

Now on a reasonable assumption that $p(E_2/D) > p(E_2)$, *i.e.* that the relevant sign, symptom, *etc.* is commoner among those with the disease in question than in the population at large, we can be sure that $\frac{p(E_2/D)}{p(E_2)} > 1$. Correspondingly we can be sure that

$$(2) \qquad \frac{p(E_1/D) \times p(E_2/D)}{(pE_1) \times p(E_2)} > \frac{p(E_1/D)}{p(E_1)}$$

So if we identify conditions under which (1) is deducible from (2), we thereby identify conditions under which E_2 corroborates E_1 in supporting the inference to D. Now an obvious condition under which (1) is deducible from (2) is when E_1 and E_2 are independent of one another in the sense of being neither positively nor negatively relevant to one another, *i.e.* $p(E_2/E_1) = p(E_2)$ and $p(E_2/E_1 \,\&\, D) = p(E_2/D)$. But is it plausible to suppose that this is at all commonly the case? Compare Jacquez and Norusis [1976]. To the extent that people don't have any symptoms or signs of ill health unless they have some disease we should certainly expect that, when E_1 and E_2 are capable of co-occurrence as symptoms or signs of the same disease, $p(E_2/E_1) > p(E_2)$. The very existence of a disease, D, that can produce both E_1 and E_2 is bound to generate this inequality. Therefore, the objection would run, we should not normally expect to find the kind of independence under which, according to the Bayesian theory—*e.g.* Lusted [1968], p. 17, *cf.* Tversky and Kahneman [1977], p. 189—corroboration takes place. Yet if that theory can give no account of corroboration it is worthless as a reconstruction of the reasoning operative in medical diagnosis, since it is a universally accepted feature of diagnosis that a suitable combination of supporting evidential facts gives added strength to the conclusion.

But the objection is less powerful than it seems. It rests on supposing that the specified independence condition is necessary, as well as sufficient, for corroboration to take place, and in fact this is not so. There are other, more appropriate conditions which are also sufficient to entail corroboration. Though Bayesian calculations do not normally rely on the existence of these conditions, the Bayesian account can be rescued by recognising their significance. Specifically, it can be proved that $p(D/E_1 \,\&\, E_2) > p(D/E_1)$ if the following six conditions hold good:

$$(3) \qquad p(D/E_1) > p(D)$$

$$(4) \qquad p(D/E_2) > p(D)$$

(5) $$p(E_1 \,\&\, E_2) > 0$$

(6) $$1 > p(D/E_1)$$

(7) $$p(E_2/D) \leqslant p(E_2/E_1 \,\&\, D)$$

(8) $$p(E_2/\bar{D}) \geqslant p(E_2/E_1 \,\&\, \bar{D}).$$

A formal proof is given in Cohen [1977].

Now of these six conditions (3)–(8) the first four would in any case be present in any normal diagnostic situation where corroboration takes place, and a moment's reflection should suffice for anyone to see that the last two conditions are scarcely an unreasonable requirement. What (7) asserts is that, given the *presence* of the disease, the existence of the symptom, sign or other evidential fact E_2 is not rendered *less* probable by the existence of E_1. After all, in a disease where two symptoms characteristically *exclude* one another, their co-occurrence can hardly *reinforce* diagnosis of the disease in question. What (8) asserts is that, given the *absence* of the disease, the existence of E_2 is not rendered *more* probable by the existence of E_1. After all, where one symptom is particularly inclined to accompany the other when the disease is *absent*, co-occurrence of the two symptoms can again hardly *reinforce* diagnosis of the disease in question.

In fact these two inequality conditions are precisely analogous to those under which the testimony of one witness corroborates that of another, or two pieces of circumstantial evidence converge to raise the probability of a suspect's guilt: *cf.* Cohen [1977], p. 101 ff. Here 'D' can be interpreted as the point at issue, 'E_1' as the first witness's assertion of this point, and 'E_2' as the second witness's assertion of it. One witness should not be more inclined to contradict the other on those occasions when the latter's testimony happens to be true, since agreement of the two testimonies would then tend to indicate their incorrectness. So, if one witness is to corroborate another, the former's inclination to give true testimony must not be reduced, but may be increased, when the latter's testimony is true. Equally, one witness must not be more inclined to give false testimony about the point at issue when the other witness does so: for example, there must be no conspiracy to deceive. Thus, both (7) and (8) express conditions that would properly be required to hold if the testimony of two witnesses to the truth of the same proposition is to be better than that of just one. On the other hand, if E_2 had to be independent of E_1, in the sense of E_1's being neither positively nor negatively relevant to E_2, many apparent cases of testimonial corroboration could not count as such. For example, since most people are honest most of the time $p(E_2/E_1)$ is surely greater than $p(E_2)$ in most cases. Hence *this* kind of independence is just as inappropriate a requirement for testimonial as for diagnostic corroboration. The formalism is right not to include it.

Similar considerations apply also to the convergence of circumstantial evidence, as when it is shown that a murder-suspect had both motive and opportunity. But if one looks for a situation in which an independence condition of this kind *does* underwrite corroboration—*i.e.* a situation where $p(E_2/E_1) = p(E_2)$ and $p(E_2/E_1 \,\&\, D) = p(E_2/D)$—one can apparently find it sometimes in games of chance. For example, let 'D' be interpreted as 'the die falls 6 uppermost'. 'E_1' as 'the die falls with a number > 2 uppermost' and 'E_2' as' the die falls with an even number uppermost'. Here $p(E_2/E_1) = p(E_2) = \frac{1}{2}$, $p(E_2/E_1 \,\&\, D) = p(E_2/D) = 1$, $p(D,\ E_1 \,\&\, E_2) = \frac{1}{2}$ and $p(D/E_1) = \frac{1}{4}$.

The inference to be drawn is that, pace Hume [1739] book I, part III, section XII, and many others, games of chance are a rather poor model for probabilistic reasoning about human conduct and physical causation. Corroboration, or evidential convergence, has an importantly different structure in regard to games of chance from what it has in regard to empirically based probability-judgments.

Of course, what emerges here is an analysis of corroboration that is employable by any account of diagnostic reasoning in terms of mathematical probabilities. It is not a specifically Bayesian analysis. But it is quite consistent with Bayesian principles, and is a satisfactory rejoinder to the objection about lack of independence between symptoms, signs, *etc.* E_2 will not normally corroborate E_1 at all unless conditions (7) and (8) are satisfied alongside (3)–(6), whatever the nature of the calculations that are used to evaluate $p(D/E_1 \,\&\, E_2)$. Moreover, a Bayesian analysis does help to bring out the nature of the conditions that affect the degree of corroboration which takes place. The analysis shows that the degree of corroboration in fact increases as

$$(9) \qquad \frac{p(E_2/E_1 \,\&\, D)}{p(E_2/E_1)}$$

increases. The rationale of this lies in a medically familiar fact: *cf.*, *e.g.* Jacquez [1964], p. 31 f. We have to distinguish genuinely additional evidence from mere redundancy, wherever disease-manifestations tend to occur together, depending on the nature of the causal linkage. Where the manifestations are closely connected together, irrespective of whether or not the disease itself is present, the level of (9) will be low, because $p(E_2/E_1)$ will be high relative to $p(E_2/E_1 \,\&\, D)$, and corroboration will be correspondingly slight. But, where the manifestations are connected together only through the causal mechanism of the disease itself, the level of (9) will be relatively high, because $p(E_2/E_1)$ will be relatively low (*i.e.* not much higher than $p(E_2)$), and correspondingly the degree of corroboration will be greater.

A fourth kind of objection to a Bayesian approach runs as follows. The clinician needs statistical data that are relativised to the same relevant demographic and clinical attributes as the patient has. But groups with just this combination of attributes would in many cases be too small for anyone to be able to make accurate estimates of the probabilities in question by reliable procedures. Differences in nosological classification between one time or place and another tend to compound this difficulty, and the difficulty is in any case obviously greater where the disease is a rare one or has bizarre complications or the patient suffers from a chance combination of diseases. And, on the other hand, where adequate statistical data *are* available for estimating $p(E/D)$, $p(D)$ and $p(E)$, one might reasonably expect them to be available also for estimating $p(D/E)$. So, even where an accurate Bayesian measurement seems possible, it could well be superfluous. This is not a general argument against Bayesianism, since for some A and B, $p(B/A)$ may be much easier to estimate than $p(A/B)$—for example, where A is a scientific generalisation and B reports the successful result of a particular experimental test on it. But, if we know enough to estimate the relative frequency of certain symptoms, signs, *etc.* among those who undoubtedly have a certain disease (presumably because we have eventually identified the latter by chemical tests, autopsies, *etc.*), we presumably also know enough to estimate the relative frequency of those who undoubtedly have a certain disease among those who exhibit the appropriate symptoms, signs, *etc.*

Perhaps a Bayesian may be tempted to rejoin that a reasonably unbiased sample of people with a certain disease may be expected to become known to members of the medical profession, since even those without appropriate symptoms or signs will eventually be revealed by autopsies, whereas an unbiased sample of people with certain symptoms or signs is more difficult to find since those who have these but no corresponding disease tend to disappear from medical view. But this would be a poor rejoinder because the strict Bayesian needs in any case to determine the value of $p(E)$. Also there is an inherent risk of bias in the fact that $p(E/D)$ would normally be determined from statistics of patients referred to hospitals while $p(D/E)$ needs to be evaluated for the population at large including those whose illnesses are sufficiently mild not to require hospitalisation.

A more characteristic rejoinder would be to deny altogether that a Bayesian account of medical diagnosis assumes the availability of adequate statistical data. If a personalist, or subjectivist, conception of probability is coupled with a Bayesian strategy for calculating with probabilities, the difficulties about statistical data are completely side-stepped. The physician who evaluates $p(D/E)$ is now supposed to be doing so in virtue of the several betting-odds he would be inclined to accept, on the basis of all his experience prior to examining the patient, for the patient's having the disease D, for the patient's exhibiting the symptoms, signs *etc.* E, and for the patient's exhibiting E given D, where the physician's overall distribution of bets is sufficiently coherent—*cf.* Kyburg and Smokler [1964], p. 11. For example, to accept betting odds of 4 to 1 on the patient's having the disease D is supposed to be tantamount to accepting a value of $\frac{1}{5}$ for $p(D)$.

But what is meant by saying that the physician's previous experience is the 'basis' for his selection of betting-odds here? This selection cannot be said to be *justified* by previous experience without reintroducing all the above-mentioned difficulties about inadequate statistical data. So apparently the physician's selection of betting-odds is 'based on' his previous experience only in the sense that the former is the causal product of the latter. And patients may well be a little concerned if their physicians can put up no rational defence when the probability of a diagnostic conclusion is challenged. Certainly it will not do to say here, as is sometimes said by personalists in other fields, that in the long run, with adequate exchange of information, the betting-odds of all rational men must coincide. For a patient may well need diagnosis and treatment within a very short run. Nevertheless, a physician's opinions are in practice often respected just because of his reputation for judicious diagnosis; and a personalist Bayesian might claim that this illustrates how we have to fall back on private, inarticulate, 'black-box' processing by a particular good brain wherever rationally defensible estimates cannot be founded on publicly available statistics. Thus the appeal to an expert's clinical judgment, which is so familiar in actual practice, can be given a theoretical representation within the personalist version of Bayesianism. But the validity of this personalist representation will be restricted to those contexts where the actual existence of the disease can in principle be determined in some other way than by reference to the symptoms, signs, *etc.* that constitute the evidence for the diagnosis. A rational bettor does not make bets that cannot be settled. So a personalist Bayesian has to assume that some eventual revelation of the truth—whether by autopsy or by surgery or by paraclinical tests—could in principle take place at some time after the observations on which he has implicitly based his own bet.

A fourth objection to the Bayesian account is more powerful, however, because it affects any method of evaluating the strength of a medical diagnosis in terms of mathematical probability. The objection concerns the importance attachable, at least from the patient's point of view, to the weight of evidence. For example, a physician may attach 95 per cent probability to the patient's having D on the evidence that he in fact manifests E_1. Then, if the physician's judgments are good ones, and a 95 per cent success-rate in diagnosis is all that his conscience, his patients, his profession or his employers require from him, he seems to be justified in diagnosing D from E_1. But more tests may reveal more evidence, E_2, and perhaps $p(D/E_1 \& E_2)$ is a great deal less than 95 per cent. So somehow or other the evaluation of a diagnosis has to be related primarily to the extent of the relevant evidence that has actually been obtained. A thoughtful clinician recognises that his reasoning may sometimes be incorrect and that his rate of error may approximate, say, 5 per cent, but he doesn't deliberately bound the thoroughness of his diagnostic examinations by the complacent acceptance of a particular rate of error. A scheme of diagnostic strategy has to be geared to the management of the unusual or atypical situation as well as the usual one, to the 5 per cent of cases that are abnormal as well as to the 95 per cent that are normal. A diagnostician at least *aims* at discovering the truth in each and every case, even if he knows that he can hardly expect actually to discover it in each and every case.

The force of this objection may be brought out by an analogy with legal justice. The criminal courts too can hardly expect to pass correct verdicts in every case. Indeed they would do very well if 95 per cent of their verdicts were correct: there is always the risk of missing or perjured witnesses, incompetent advocates, perverse or corrupt juries, *etc.* So suppose the only evidence presented to the court in the trial of a man accused of possessing illegal drugs is that he possesses a rainbow-painted car and that 95 per cent of those who possess rainbow-painted cars also possess illegal drugs. Would it be justice to condemn him? Should he be deprived of his liberty? No. But why not? Because there isn't *enough* evidence. Further enquiry might show that there was absolutely no trace of illegal drugs anywhere in his house or car. Correspondingly we can't suppose a criminal court to incorporate its diagnostic ambition—the ambition to be right in 95 per cent of its cases—within the standard of proof that it adopts. For, if all that was required was a 95 per cent probability, the man with the rainbow-painted car would be justly condemned. Instead the courts must aim at a correct verdict in every case, and that means that they need to be presented with as much as possible of the relevant facts. If the accused is to be condemned, his guilt must be proved beyond reasonable doubt. So the trier of fact in a criminal court is in a situation that is analogous in some respects to that of a diagnostic physician (even though medical diagnosis does not presume that a patient is innocent of a particular disease unless his having it is proved beyond reasonable doubt). The trier of fact, like the physician, cannot base his verdict *merely* on a high mathematical probability. No doubt the measurement of mathematical probability is sensitive to the impact of corroborative or converging evidence, as we have already seen. But without looking to the extent—or 'weight', as Keynes [1921], p. 71 ff. called it—of the evidence that we do at present have we cannot tell how much significance to attach to the fact that it happens to converge towards a very high mathematical probability for such-or-such a conclusion.

3 Let us now look at the main theory of medical diagnosis that is offered as a rival to Bayesian modes of evaluation: *cf.* especially Kleinmuntz [1965] and Feinstein [1973]. We shall see that this theory has a complementary virtue, but also a complementary vice. It has the virtue of making the ordinal or comparative grading of evidential weight quite easy. But it assumes, rather optimistically, that in the end the evidence would normally be complete and wholly certain diagnoses would normally become available, even though considerations of time or money may in practice often force the process of collecting evidence about a particular patient to stop before it is complete.

According to this theory diagnostic reasoning may be represented either by a decision-table or by an equivalent flow-chart. In the decision-table representation a series of questions are tabulated, each being susceptible of a yes-or-no answer. An appropriate diagnostic conclusion is then assigned to each familiarly encountered combination of 'yes' and 'no' answers. A flow-chart representation carries the same total amount of information, but may economise in the amount of symbolism that is needed to express it. Its non-terminal boxes are labelled with questions, and two arrows—one marked 'yes' and the other 'no'—lead off from each such box to other boxes: the terminal boxes are labelled with appropriate diagnostic conclusions. So a flow-chart representation can economise with symbolism if several different combinations of answers justify the same conclusion or if one particular answer to one of the questions recurs in the evidence for a number of different diagnoses. A decision-table gives us a conspectus of the detailed variety of obtainable information, with corresponding conclusions, while a flow-chart represents best the variety of different possible sequences of reasoning. A single vast decision-table, or flow-chart, could represent the total diagnostic potential of current medical science, with every stage of reasoning made fully explicit. But in practice one might suppose that a number of much shorter and more specialised diagrams would have greater utility, such as the flow-chart for examining part of a barium enema that was given in Medical News [1966], p. 36, or the hundred or more flow-charts that have been published in *Patient Care* in recent years.

In practice, because of the risks involved in delay, therapy of some kind may have to be commenced long before all the required information is available and thus before correct therapy is uniquely determined. Or perhaps it would be prohibitively expensive to carry out all the laboratory tests that are required to provide ideally complete information in a particular case. Unless there are special risks in doing so, the obvious course then is to select the diagnosis that seems most probable on the available evidence. Equally obviously, the merit of this selection depends not only on the relevant degree of mathematical probability, but also on the actual extent to which the questions tabulated on the decision-table have been answered, and thus on the extent to which the relevant route on the flow-chart has been followed through. Measures of probability are familiar; but how do we evaluate the merit of a diagnostic conclusion in terms of the extent of available information?

Writers on medical diagnosis are normally more interested in the heuristic aspect of the reasoning involved here than in its probative aspect. That is to say, they are more interested in solving the methodological problem 'What is the rational way to discover a patient's disease?' than in solving the logical problem 'What is the rational way to justify a particular diagnosis?' But the problems are intrinsically interconnected. A diagnosis reached in the methodologically correct way is justified thereby, and an evaluation of the extent to which a particular diagnosis is justified is *ipso facto* an evaluation of the extent to which a patient's disease has indeed been discovered. So it is of methodological as well as logical interest to determine the exact nature of the mode of evaluation that is appropriate to decision-table or flow-chart representations of diagnostic reasoning.

My contention here will be that this mode of evaluation is essentially the same as that which Cohen [1977], p. 121 ff. calls inductive, or Baconian, probability. The point is that the reliability of a generalisation, in experimental science as also elsewhere, is best graded, as Bacon [1620] saw, by the variety of potentially relevant circumstances that fail to falsify it. (Relevance in this context is established by the successful falsification of generalisations about kindred topics.) The clearest examples of such Baconian gradation are to be found where the reliability of a generalisation is gauged by the results of a controlled experiment and is taken to be higher or lower in accordance with the number of relevant conditions that are controlled in the (successful) experiment. For instance, the curative value of a drug might be established by experiments in which patient's medical histories, ages, diets, *etc.*, were all controlled. The more comprehensive the test that a generalisation survives, the greater the reliability that is ascribable to it. Bacon rightly thought of such tests as putting general questions to nature, so that as the interrogation increases in thoroughness its outcome increases in certainty. It follows that, if a generalisation would be falsified under conditions in which the circumstance V_1 were present, it may be protected against such falsification and thus achieve a higher grade of reliability if it is qualified accordingly. That is, in appropriate cases 'Anything that is A and V'_1 is also B' will be more reliable than 'Anything that is A is also B', because the presence of V'_1 excludes that of V_1. (*E.g.* we might ascribe curative value to the drug only when the patients are young.) Correspondingly an inference to a thing's being B from the premise that it is both A and V'_1 is stronger (or weightier) than an inference to that conclusion from a premise stating only that it is A. Similarly an inference from the premise that it is both A and V'_1 and V'_2 may be stronger still. Consider then those pathological generalisations that assert any person to have such-or-such a disease if he has such-or-such symptoms, signs, biographical circumstances and/or paraclinical test-results. Clearly at least one way of grading the reliability of such generalisations is the Baconian method that has just been described. And correspondingly, if the basic research has been done satisfactorily and the textbook generalisations contain all the right qualifications, a diagnosis that is made in conformity with these generalisations will vary in strength (or weight) according as the evidential premisses of the diagnosis for a particular patient include more and more of the circumstances that are mentioned in the antecedent clause of the appropriate generalisation.

Ideally we should be able thus to rank-order any given set of diagnoses within a particular field of medicine. In order to guarantee comparability of strength in every case, however, between one diagnosis and another within a particular field of medicine, we need to assume a uniform sequential ordering for the process of collecting evidential data. Such an ordering is implicit in the decision-table or flow-chart representation of diagnostic reasoning. It is also a standard requirement for the list of relevant variables that is to be manipulated in controlled experiments if comparative grades of Baconian reliability are to be established by these experiments. Obviously, there is much to be gained by varying just the most important factors—*i.e.* the circumstances with greatest falsificatory potential—if shortage of time or resources allows only some factors to be varied. The quicker we can get false hypotheses out of the way, the better. So in an idealised reconstruction of diagnostic reasoning, as of experimental research, one would naturally expect an ordering in which the earlier stages eliminate more hypotheses than the later ones. Disease-manifestations, as Jacquez [1964], p. 31 f. put it, need to be given a hierarchy of values. As a result the different kinds of data that are in fact collected for diagnostic purposes in a particular area of medicine standardly constitute a Guttman scale, in the sense that, if we know how many different kinds of data are present in the evidence actually obtained for a particular diagnosis, we should be able to name which kinds they are.

No doubt practical difficulties may often intervene to prevent adoption of the theoretically optimal ordering. But we must then be prepared to accept consequential restrictions on the comparability of results. These constraints are not an argument against accepting a Baconian *logic* for judgments of diagnostic weight. They merely affect the extent to which a flow-chart can give a uniformly applicable account of diagnostic *methodology*.

Now, if it is correct thus to interpret the strength or weight of a diagnosis as a Baconian probability, when the diagnosis is carried out in accordance with the decision-table or flow-chart model, we can use what is known about the logic of Baconian probability in order to derive corresponding properties for this concept of diagnostic weight. In that way we can clarify what constraints one such evaluation of weight imposes on another, just as the mathematical calculus of chance articulates what constraints one Bayesian evaluation of probability imposes on another. And since a generalisation of the Lewis–Barcan modal system S4 articulates the logical syntax of statements about Baconian probability, as in Cohen [1977], p. 229 ff., we shall find there also a logic of statements about diagnostic weight.

For example, a conjunction of diagnoses acquires as much weight from given evidence as either diagnosis has separately, if they have equal weight, or as the less weighty one has, if they are unequal. So the diagnostic output of a physician does not lose its overall strength as he sees more and more patients. And obviously this is a feature that distinguishes the weight of diagnostic statements very sharply from their mathematical probability, since the multiplicative principle for the probability of a conjunction must entail some reduction in the probability of the overall correctness of a physician's diagnostic output, as he sees more and more patients. It will be less and less *probable* that he got the diagnosis right in every case, even when each patient has precisely the same symptoms, signs *etc.* Yet there is also a sense in which, if every constituent diagnosis is as solid as the evidence permits, then the diagnostician's total achievement is also just as solid. This sense is captured by the Baconian analysis of diagnostic weight.

Again, if the evidence is reliable and gives weight to a particular diagnosis, though the diagnosis is not conclusively proved, a Baconian analysis requires us to say that no weight at all is given to the denial of the diagnosis. And here again there is an important disanalogy with the logic of probability-judgments, since if the probability of a diagnosis falls short of complete certainty, the probability of its negation must be greater than zero. The point is that if the balance of evidence builds up in favour of diagnosing that the patient has disease *D*, the weight of this build-up may be greater or less, depending on how much relevant evidence has been collected, but since the balance of evidence favours the presence of *D* it does not favour the absence of *D* and a zero-grade weighting has therefore to be given to any denial that *D* is present.

So far as evaluations of diagnostic weight embody judgments of Baconian probability, their logic cannot be mapped in any seriously viable way on to the mathematical calculus of chance. It is not just that the two concepts obey different principles for conjunction, negation, *etc.* The logical gulf between the two concepts is much wider because no function can be constructed out of the resources of the calculus of chance that will properly evaluate the strength or weight of a diagnosis that has been carried out in accordance with the flow-chart or decision-table theory.

This would be true even if strength or weight here were an additive property, as is shown in Cohen [1977], pp. 188 ff. and 224 f. But in fact we cannot treat it as an additive property, because the reliability-grade of the pathological hypothesis from which such a diagnosis derives depends on the power of the hypothesis to resist falsification by cumulative combinations of relevant factors, and this power is not an additive property either. Many pairs of casual factors, for example, have an explosive potential when combined with one another, though they are only of slight significance when occurring apart. Cigarette-glow and petrol-vapour constitute a familiar example of this in everday life, or alcohol and barbiturates in medicine. So a Baconian weight-function cannot quantify the strength of a diagnosis, in the way that a successful Bayesian investigation would quantify its probability. Diagnostic weight submits at best to rank-ordering. The resistance that clinicians sometimes feel to quantifying the strength of their diagnostic convictions has a solid rationale. It may be that some concepts of weight admit a Bayesian analysis. But diagnostic weight is not one of these.

4 What emerges, therefore, is that a Bayesian and a flow-chart (or decision-table) evaluation ought not to conflict with one another, because when properly constructed and appropriately applied they are concerned with different issues and complement one another. The Bayesian evaluation, or indeed any evaluation in terms of mathematical probabilities, declares how far the evidence favours, and how far it opposes, the suggested diagnosis, *on the assumption that this is all the available evidence.* The flow-chart evaluation, or indeed any evaluation in terms of evidental weight or Baconian probability, reveals just *whether* the balance of the evidence obtained favours the suggested mathematical probability, or not, and if so *how much* of the potentially available relevant evidence is constituted by the evidence actually obtained. In practice the desirability of having any kind of quantitative evaluation in terms of mathematical probability, as a final outcome of the diagnosis, tends to decline as the Baconian probability, or weight, increases. Where there is no time, perhaps, or no inclination to carry out a thorough examination, the patient may have to be treated on the basis of a positively relevant mathematical probability for anyone with his immediately obvious signs or symptoms. At the other extreme, where a thorough and exhaustive collection of relevant evidential data can be made, including the results of appropriate paraclinical tests, the patient's condition may be so well understood that an evaluation in terms of mathematical probability would be superfluous, because so much of the causal process at work has been established that the probability would be only a negligible distance short of 1 (if indeed any worthwhile estimate of the probability were obtainable for so specific a combination of circumstances). But in between these two extremes there is plenty of room for combining both modes of evaluation. To the extent that further evidential data, which might produce certainty, are still lacking, it may well be necessary to add a judgment of mathematical probability to that of evidential weight. But the significance of such a judgment cannot be properly appreciated without due consideration's being given to the weight of the evidence on which it is based. And this applies not only to straightforward judgments of mathematical probability but also to judgments that evaluate some function of the relevant probabilities like Mycin's 'certainty factor'—*cf.* Shortliffe [1976], p. 171. In general it is in the patient's interest to require as weighty a diagnosis as the risks of obstructing therapy or provoking complications allow, and to have his physician rely as little as possible on

the calculation of mathematical probabilities. Economies of finance and professional effort, however, are obviously best served by the opposite policy.

It follows that computerisation may promote diagnostic efficiency in two rather different ways. So far as Bayesian evaluations are concerned, it is clear that a programme for carrying out the requisite calculations on a computer can save time and effort, where the requisite statistical data are available: *cf. e.g.* Overall and Williams [1963]. So far as Baconian evaluations are concerned, however, it is also clear that programmes can be constructed for carrying out that part of a routine diagnostic enquiry which consists in the collection of information about a patient's symptoms, life-style, environment, *etc.* or in monitoring the patterns revealed by paraclinical tests. Such a programme can ask the relevant questions and absorb the relevant answers in a reliably impersonal manner, with appropriate feedback to the questioning process—*cf. e.g.* Kleinmuntz [1972] and Garry [1976]—though there are many problems here that are yet to be overcome, as Schaffner (forthcoming) makes clear. These two kinds of computerisation are just as little rivals to one another (except perhaps for the use of computer time) as are the two different types of diagnostic evaluation.

Finally, one needs to beware of drawing the wrong conclusions from the fact that in a particular series of trials a Bayesian programme may have been as successful in making the right diagnosis as a group of well-qualified clinicians: *cf.* Overall and Williams [1963]. This is exactly what one should expect if the computer is really supplied with just the same *weight* of evidence as the clinicians consciously or unconsciously obtain. It is quite another matter to judge in a particular case, whether that weight of evidence is heavy or light, adequate or inadequate; and for this we need Baconian, not Bayesian reasoning.[1]

The Queen's College,
Oxford University

REFERENCES

Bacon, Francis [1620]: *Novum Organum*, London.
Cohen, L. J. [1977]: *The Probable and the Provable*, Oxford.
Feinstein, A. R. [1973]: 'An Analysis of Diagnostic Reasoning', *Yale Journal of Biology and Medicine* 46, 1973, pp. 212–32 and 264, 283, and 47, 1974, pp. 5–32.

[1] I am grateful for some helpful comments on an earlier version of this paper to Jonathan E. Adler, Godfrey Fowler, Kenneth F. Schaffner, and R. G. A. Williams, and to participants at a meeting of the Society for Exact Philosophy in Montreal, Quebec, on June 5, 1979, and at a session of the 6th International Congress for Logic, Methodology and Philosophy of Science at Hannover, West Germany, on Thursday, August 23, 1979.

FEINSTEIN, A. R. [1977]: 'Clinical biostatics XXXIX. The haze of Bayes, the aerial palaces of decision analysis and the computerized Ouija board', *Clinical Pharmacology and Therapeutics* **21**, pp. 482–96.

GORRY, G. A. [1976]: 'Knowledge-Based Systems for Clinical Problem Solving' in F. T. de Dombal and F. Grémy (*eds.*): *Decision Making and Medical Care: Can Information Science Help?* Amsterdam, pp. 23–31.

HUME, D. [1739]: *A Treatise on Human Nature*, London.

JACQUEZ, J. A. [1964]: 'The Diagnostic Process: Problems and Perspectives' in J. A. Jacquez (*ed.*): *The Diagnostic Process*, pp. 23–37.

JACQUEZ, J. A. and NORUSIS, H. J. [1976]: 'The importance of symptom non-independence in diagnosis', in F. T. de Dombal and F. Grémy (*eds.*), *Decision Making and Medical Care: Can Information Science Help?* Amsterdam, pp. 379–92.

KEYNES, J. M. [1921]: *A Treatise on Probability*. London.

KLEINMUNTZ, B. [1968]: 'Diagnostic problem-solving by computer', *Japanese Psychological Research*, **7**, pp. 189–94.

KLEINMUNTZ, B. [1972]: 'Medical Information Processing by Computer', in J. A. Jacquez (*ed.*) *Computer Diagnosis and Diagnostic Methods*. Springfield, pp. 45–72.

KYBURG, H. E. Jr. and SMOKLER, H. C. [1964]: Editorial introduction to *Studies in Subjective Probability*, New York.

LANGE, H. J. [1976]: 'Decision-making by Doctors and aided by Computers', in F. T. de Dombal and F. Grémy (*eds.*), *Decision Making and Medical Care: Can Information Science Help?*, Amsterdam, pp. 11–22.

LEDLEY, R. S. and LUSTED, L. B. [1959]: 'Reasoning Foundation of Medical Diagnosis', *Science*, **130**, 3 July 1959, pp. 9–21.

LUSTED, L. B. [1968]: *Introduction to Medical Decision Making*, Springfield.

MEDICAL NEWS [1966]: Medical News section, *Journal of American Medical Association*, **19612**, pp. 1–42.

OVERALL, J. F. and WILLIAMS, C. M. [1963]: 'Conditional Probability Program for Diagnosis of Thyroid Function', *Journal of the American Medical Association*, **183**, pp. 307–13.

SARDEMAN, W. A. [1964]: 'The history and physical examination', in J. A. Jacquez (*ed.*) *The Diagnostic Process*, University of Michigan Press, 1964.

SCHAFFNER, K. F. (forthcoming): 'Problems in Computer Diagnosis', forthcoming in *Ethics and Clinical Diagnosis* in the *Philosophy and Medicine* Series, (*eds.*) H. T. Engelhardt and S. Spicker.

SHORTLIFFE, E. H. [1976]: *Computer-Based Medical Consultations: MYCIN*, New York: Elsevier.

TVERSKY, A. and KAHNEMAN, D. [1977]: 'Causal Thinking in Judgment under Uncertainty', in R. Butts and J. Hintikka (*eds.*), *Basic Problems in Methodology and Linguistics*, Dordrecht, 1977, pp. 167–90.

WULFF, H. R. [1976]: *Rational Diagnosis and Treatment*, Oxford: Blackwell.

Discussion

POPPERIAN LANGUAGE-ACQUISITION UNDEFEATED

After exposure to a finite (though large) body of data exemplifying the language of their elders, children regularly succeed in learning the language, in the sense that they go on in due course to produce utterances distinct from any of those that they have heard but which are nevertheless well-formed in their elders' language (and they refrain from uttering ill-formed word-sequences). This achievement is formally akin to that of an adult scientist who, from a finite data-set, infers a theory which permits him to make successful predictions about cases which he has not observed. Linguists commonly say that the child infers a 'grammar' from his data, and that the child's 'grammar' governs his subsequent linguistic behaviour. Linguists think of a 'grammar' as the same kind of entity as a scientific theory, except that the possessor of a grammar is not (or at least is not necessarily) consciously aware of its content. There are large conceptual problems in the notion of a man's behaviour being governed by a 'theory', no aspect of which may have ever entered his conscious awareness and which he may be wholly incapable of stating to an enquirer, but I believe that these problems are not important for the matters discussed below.

Noam Chomsky has argued in many of his writings that the invariable success of children faced with the task of first-language-acquisition can be explained only by supposing that humans inherit a language-faculty which builds much of the structure of a grammar into the individual's mind from the beginning, so that only limited details remain to be decided from the data of experience. He supports this by arguing that all individuals formulate grammars sharing certain common structural properties, although it is unreasonable to suppose that only grammars having those properties are logically compatible with the data-sets available to individual language-learners, and grammars possessing the properties in question have no advantage in terms of general methodological criteria (such as simplicity) over potential grammars which lack them. (The claim that different individuals' grammars share these properties, which have to do with the hierarchicality of syntactic structure, is in turn supported by the claim that the different languages of the world can be observed to share them, an observation which is surprising inasmuch as we have no *a priori* reason to expect that efficient communication-systems would need to possess these properties.)

I argue in my [1978], [1980] that Chomsky's strongly anti-empiricist account of language-acquisition is ill-founded, and that we should instead think of a child formulating a grammar on the analogy of an adult scientist formulating an explicit theory, as viewed by the Popperian philosophy of science. That is, the individual child begins with no reliable knowledge but builds up a grammar by formulating and testing a series of fallible conjectures, choosing between alternative hypotheses compatible with his data by reference to general methodological criteria (notably, strength and simplicity). Chomsky's hierarchical syntax-universals are not evidence for an innate grammar, because, using ideas developed by Herbert Simon ([1962]), I am able to show that just these properties are a (non-obvious) logical consequence of the fact that grammars are evolved step by step (as the Popperian analogy requires) rather than emerging almost full-blown from the start (as Chomsky argues). (Certain of Chomsky's alleged linguistic universals are not predictable via Simon's argument, but these are shown not in fact to be universal.)

According to Stephen Stich ([1979]) I fail to undermine Chomsky's anti-empiricism, because my explanation of the hierarchical syntax-universals (an explanation which, at least for the sake of argument, Stich accepts) does only part of the job which faces anyone seeking to account for the language-learner's achievement. I explain why a child invariably infers from his data a 'hierarchical' rather than a 'non-hierarchical' grammar; but, Stich argues, what needs to be explained is why a child invariably infers the *correct* 'hierarchical' grammar. Just as there will always be alternative theories compatible with any finite body of data, so there will always be alternative grammars *conforming to Chomsky's linguistic universals* and compatible with a child's data. A rational scientist could not (except by a lucky chance) get the right answer on the basis of the experience available to an individual child; therefore the fact that children always do formulate a grammar identical with that of their elders (at least with respect to their implications for observable linguistic behaviour) shows that children must be innately equipped with some more powerful theory-finding machinery than is used by a rational scientist (Stich [1979], §IB).

To my mind the wealth of linguistic data available to a child, the tendency to choose strong and simple theories which can be observed to guide the theorising of adult scientists (without them necessarily reflecting consciously on these considerations) and which can therefore plausibly be attributed also to the child, and the constraints implied by Simon's argument, will between them narrow down the set of possible grammars to a class whose members will differ from one another rather little (although they may be numerous, perhaps infinitely so). But I do not need to establish this in order to counter Stich; I can allow, as he suggests, that the factors just listed may conjointly be compatible with grammars which differ wildly with respect not merely to their internal structure but to the predictions they make about observable behaviour. All I need to do, essentially, in order to defeat Stich is to ask: *How does he know* that children invariably 'get the right answer' when learning their elders' language?

The only way one *could* know this would be to act as a rational scientist by formulating theories about the grammars of a number of individual members of a language-community, in order then to show that these grammars were very similar to one another, or to the grammars of members of the previous generation (which would also have to be discovered by acting as a rational scientist). Now Stich may play the game whichever way he wishes. If he says that the rational scientist will have no means of choosing between wildly different grammars on the basis of the kind of data-set available to a child, then I reply that the rational scientist will in that case certainly not be in a position to state that the grammars of two individuals are near-identical, since no data are available to the rational scientist that would not be available equally to a child. If, on the other hand, Stich says that the rational scientist *can* ascertain the nature of individuals' grammars precisely enough to confirm that they are closely similar, then I reply that in that case the child can ascertain the precise nature of his elders' grammar by the same techniques.

The advantages would appear to lie on the side of the child: he only has one job of grammar-discovery to carry out, while the adult scientist seeking evidence in favour of Stich's thesis has to formulate independent theories about the grammars of a number of individuals in order to show that the differences between these grammars are, regularly, small. Furthermore, the motive for an emotionally normal child to acquire his elders' language is presumably overwhelming, while academic research is only one interest among others in an adult scientist's life; and few adult linguistics researchers are provided with such excellent research facilities and such immediate and powerful reinforcement for advances in their research as are routinely provided by parents for their language-learning offspring.

Perhaps Stich would meet this objection as follows. If the rational scientist chooses to investigate his own language-community, he begins with a sort of privileged access to the grammar of one member of the community, namely himself. Furthermore he can ask other members of the community to report on their grammars, and this provides a very useful category of data not available to the language-learning child (who can only observe speakers' linguistic behaviour and could not understand their reports about their grammars even if they were to supply any).

But this would be circular. On the Popperian view of language-acquisition, the grammar to which the scientist may be thought to have privileged access (his own grammar) and the grammars which he may be able to find out about relatively easily by asking people rather than by doing scientific research (the grammars of other members of his own language-community) were themselves brought into being through a process logically quite akin to conscious scientific theorising. It is only if Stich and Chomsky are right in their anti-empiricist approach to language-acquisition that the special category of data available to a scientist who chooses to investigate his own language-community includes truths which might not be discoverable by a rational scientist working exclusively with the kind of behavioural data available to a child. Stich is not entitled to assume the correctness of this view in the course of an argument purporting to show that it is correct.

(In fact, the evidence available suggests that speakers' conscious beliefs are often quite wrong as reflections of the grammars which actually govern the speakers' linguistic behaviour, and that members of a single language-community differ considerably in their beliefs about their language where these beliefs concern aspects of the language that are not the subject of a shared cultural tradition of explicit teaching. For both these points, see *e.g.* Labov [1975], Snow and Meijer [1977].)

One might think that the claim that the grammars of various members of a language-community are identical is an extremely strong hypothesis, so that Stich could reasonably suggest that it deserves to be believed in the absence of counter-evidence, even if the evidence available about any one speaker is far too scanty to determine the nature of his grammar. However, if we cannot determine any one speaker's grammar from the data of behaviour then the hypothesis that various speakers' grammars are identical is actually unrefutable, because no linguistic behaviour can be 'refuted' or 'contradicted' by any other linguistic behaviour—only a *theory* can be refuted by behaviour or contradicted by another theory.

Stich thus gives us no reason to believe his claim that children are more successful than Popperian rational scientists would be at inferring the correct grammar from their data. Indeed, it is only out of deference to Stich that I have so far used the phrase 'infer the correct grammar' as if it represented an achievement that *might* meaningfully be ascribed to children—I do not believe it is. At the end of his section II, Stich complains that I overlook a disanalogy between the Popperian account of science and any possible account of language-acquisition: Popper does not claim that the scientific theory in a given domain which is favoured by our methodological rules at any given time will be correct, but a successful account of language-acquisition must explain the acquisition within a finite time of the correct grammar. I agree that there is a disanalogy, but this is not it. The disanalogy I see is that, on Popper's realist view of science, there is a true theory of any domain of natural phenomena waiting to be discovered, although we will have no way of knowing that we have discovered it if we ever do; while a language, such as English, has no existence apart from the various grammars of individual members of the language-community, so that one cannot speak of a given individual's grammar as 'correct' or 'incorrect' but only as 'relatively orthodox' v. 'relatively eccentric', or the like. One might call a child's grammar 'correct' or 'incorrect' if the child was exposed to the linguistic output of just one adult whose grammar was stable; then the child's grammar would be 'correct' if it coincided exactly with the adult's. But, in the first place, a child will commonly be exposed to many individuals' speech, exemplifying at least slightly different languages. Furthermore, it is not clear to me why Chomsky and Stich assume without argument that individuals cease to modify their grammars after a certain (fairly early) age; it seems not implausible to suppose that we go on refining our grammars in increasingly subtle respects throughout our lives. (If it can be shown that we do not, however, this will not destroy the analogy between language-acquisition and scientific advance; individual scientists also show a tendency to become less flexible in their beliefs as they grow older.)

There are hints in Stich's article that he disagrees not merely with my adaptation of Popperian ideas to the phenomenon of language-acquisition, but also with the Popperian view of scientific theorising. If so, then there is an escape available to Stich from the fork on which I have sought to impale him. That is, Stich might argue that children are highly successful in the language-acquisition task, that a Popperian rational scientist guided only by general methodological criteria of strength and simplicity could not achieve comparable success with the same data, and that we can know that children outperform Popperian rational scientists because we are not Popperian scientists ourselves—in our conscious scientific theorising (about other people's grammars, among other things), as well as in our youthful language-learning, we exploit innate theory-finding machinery more powerful than anything envisaged by Popper. (There is little doubt that this is the line Chomsky would nowadays take; see *e.g.* his [1976], pp. 24–5.)

But this is merely to argue that, if *for independent reasons* we hold a relatively rationalist theory of knowledge-acquisition, then we may need to make our account of children's language-acquisition conform to this theory. The claim of Chomsky's which I set out to undermine in my [1978] was that language-acquisition phenomena provide *evidence* for a rationalist theory of knowledge-acquisition which ought to be accepted as cogent by people who do not start with a prejudice in favour of rationalism. I also believe that strongly rationalist theories of knowledge-acquisition can be shown to be implausible with respect to issues independent of linguistics; but that is another story.

GEOFFREY SAMPSON
University of Lancaster

REFERENCES

CHOMSKY, A. N. [1976]: *Reflections on Language*. London: Temple Smith.

LABOV, W. [1975]: 'Empirical Foundations of Linguistic Theory', in R. Austerlitz (*ed.*), *The Scope of American Linguistics*. Lisse: Peter de Ridder Press.

SAMPSON, G. R. [1978]: 'Linguistic Universals as Evidence for Empiricism', *Journal of Linguistics*, **14**, pp. 183–206.

SAMPSON, G. R. [1980]: *Making Sense*. To be published by Oxford University Press.

SIMON, H. A. [1962]: 'The Architecture of Complexity', *Proceedings of the American Philosophical Society*, **106**, pp. 467–82, reprinted in *The Sciences of the Artificial*, Cambridge, Mass.: M.I.T. Press, 1969.

SNOW, C. and MEIJER, G. [1977]: 'On the Secondary Nature of Syntactic Intuitions', in S. Greenbaum (*ed.*): *Acceptability in Language*. The Hague: Mouton.

STICH, S. P. [1979]: 'Between Chomskian Rationalism and Popperian Empiricism', *British Journal for the Philosophy of Science*, **30**, pp. 329–47.

Review Article

ARMSTRONG'S THEORY OF UNIVERSALS*

'Scientific Realism' will suggest to readers of this journal a view that contrasts with scientific instrumentalism. David M. Armstrong's *Universals and Scientific Realism* is concerned with the reality of properties and relations rather than with the theoretical entities of science. Armstrong's Realism contrasts with Nominalism and is Scientific in contrast with being *a priori*. Although arguments for the reality of universals are *a priori* as arguments in metaphysics typically are, determination that a certain universal exists is never an *a priori* matter. Only science can tell us what kind of universals there are. Armstrong completely rejects the identification of universals with the meanings of general terms and thus rejects the inference from the meaningfulness of a general term to the existence of a corresponding universal. There is no neat correlation between meaningful predicates and universals. It is the resemblance between particulars or sameness of nature and the facts of causation that are unaccountable without the existence of universals, not the fact that some terms apply to more than one particular.

Short chapters, 1–24, are divided evenly between volume I, *Nominalism and Realism*, and volume II, *A Theory of Universals*. The chapters are grouped in seven parts, (1) Preliminaries, (2) Theories of Universals, and (3) Particulars, in volume I, (4) Predicates and Universals, (5) Properties and Relations, (6) The Analysis of Resemblance, and (7) Higher-Order Universals, in volume II. The results of volume I are mostly negative. Five kinds of Nominalism are distinguished. (1) Predicate, (2) Concept, (3) Class, (4) Mereological and (5) Resemblance Nominalism respectively hold that an object has a property, *F*, if and only if it (1) falls under the predicate '*F*', (2) falls under the concept *F*, (3) is a member of the class of *F*s, (4) is part of the aggregate of the *F*s, or (5) suitably resembles the paradigm cases of an *F*. Although these do not share all the same difficulties, there are some serious difficulties they all share because of their common claim that a thing's having a property is always a matter of its being related to something external to it. First, it appears on reflection that a thing could have a property even if such a thing external to it did not exist. Second, the relation to this thing needs to be accounted for but cannot be without either circularity or vicious regress. Third, the causal powers of a thing are determined by its properties, and these powers appear to be due to something intrinsic to the thing rather than to something beyond it. Difficulties of this sort also attend varieties of Realism which hold that something has a property, *F*, if and only if it has a certain relation to a property, universal, or transcendent Form. The connection or tie between a thing and a property it

---

*Review of ARMSTRONG, D. M. [1978]: *Universals and Scientific Realism*. Cambridge: Cambridge University Press. Volume I, *Nominalism and Realism*. Pp. xvi+149. £7.50. Volume II, *A Theory of Universals*. Pp. viii+190. £7.50.

has is somehow not a relation. Armstrong argues against the view of G. F. Stout and D. C. Williams that different things cannot have numerically the same property. Properties are universals. Although every property is had at some time by something, there is no logical limit to the number of things that can have a property.

There obviously is a logical limit to the number of things that be numerically identical to a given particular. There is thus no property such as *being identical to David M. Armstrong*. Neither is there a property of *being self-identical*, although there is no logical limit to the number of things to which the predicate 'is self-identical' applies. The applicability of this predicate to any thing can be determined *a priori*, and Armstrong insists that the existence of a property cannot be determined *a priori*. He also suggests that a property must endow any particular that has it with causal powers, and neither *being self-identical* nor *being identical to such-and-such* seems to endow anything with a causal power. He does not mention how his treatment of identity bears on arguments like the following: If x is identical to y, and x has property F, then y has property F. Therefore, if a is identical to b, then b has the property of *being necessarily identical to a* and the property of *being determinately* (*or definitely*) *identical to a*, since a has both these properties (Kripke [1971], Evans [1978]). These arguments fail, on Armstrong's view, because his reasons for denying that *being identical to a* is a property apply just as well to *being necessarily identical to a* and to *being determinately identical to a*. The thesis that true identity statements are necessarily true receives independent support from Kripke's theory of rigid designation, and at least some objections to the thesis are undermined by Kripke's distinction between the necessary and the *a priori*. The thesis that an identity statement must either be determinately true or be determinately false seems to me in rather worse shape once the argument from Leibniz's Law (so called) is rejected. Sorites arguments which show up the vagueness of 'tall', 'bald', 'blue' and 'wealthy' appear also to apply to identity. A series of small alterations, no one or two or three of which is sufficient to destroy a thing's identity, can be cumulatively sufficient to alter an object into a numerically distinct object.

In holding that there are universals, Armstrong rejects the Nominalist claim that there is nothing but particulars. He wants also to reject the Universalist claim that there is nothing but universals. Chapter 9, 'Are particulars reducible to universals?' bristles with arguments supporting a negative answer. (The density of short arguments is high throughout the book.) One argument against the view that particulars can be reduced to complexes or bundles of universals is directed against the Principle of the Identity of Indiscernibles which the reduction requires. It is evident to Armstrong that the Strong form of the Principle, which holds that distinct things must differ in at least some non-relational properties, is not a necessary truth. If science discovers that distinct electrons can share all their non-relational properties, it is not a truth at all. Armstrong argues that the Weak form of the Principle, which holds that distinct things must differ in at least some property, relational or non-relational, is not really weaker than the Strong form of the Principle if particulars are nothing but bundles of universals.

> The relational properties of a particular will be a matter of its having relations to other particulars of a certain sort. But the notion of a particular

> is the notion to be analyzed. So the Bundle theorist must analyze the situation where a particular has a relational property by saying that one bundle of properties has a certain relation to another bundle of properties.
>
> Every relational property that a particular has, then, will be a matter of a bundle of properties standing in a certain relation to other bundles of properties. The bundles of properties themselves will therefore have to be bundles of *non-relational* properties. Only so can there be bundles to have relations to other bundles. Each of these bundles is a different particular, so no two bundles can contain exactly the same non-relational properties. That is to say, the Bundle theory is forced to appeal to the *Strong* version of the Identity of Indiscernibles. (I, pp. 94–5.)

If *P* and *Q* are incompatible non-relational properties, the following situation appears to be possible: some bundle containing *P* is next to some bundle containing *P* but is not next to any bundle containing *Q*, and some bundle containing *P* is next to some bundle containing *P* and is also next to some bundle containing *Q*. In this situation, there are at least three bundles, two of them *P* and one of them *Q*. So far as the situation is described, two bundles distinguished from each other by relational properties have the non-relational property *P* in common. Why then cannot they have all their non-relational properties in common? I do not see that Armstrong gives any good reason they cannot.

After first reading the above argument and pausing because of the consideration just outlined, I was somewhat disarmed to read in the next paragraph that

> If the previous argument is sound, discussion of the Weak form of the Principle is irrelevant. But perhaps it is not sound. It may therefore be worth showing that even the Weak Principle is not a necessary truth. (I, p. 95.)

Although here Armstrong needs to justify his discussion of putative counterexamples to the Weak Principle formulated by Black, Ayer and Broad, the editorial comment is typical of his attitude toward his own arguments throughout the book. If they are all sound, there are many more than he needs to support his conclusions. As he candidly admits, however, some are better than others.

While rejecting attempts to reduce particulars to universals, Armstrong also rejects attempts to explain the particularity of particulars by reference to bare particulars. Universals and particulars cannot exist without each other. He holds both a Principle of Instantiation, that for each *N*-adic universal, *U*, there exists at least *N* particulars such that they are *U*, and a Strong Principle of the Rejection of Bare Particulars, that for each particular *x*, there exists at least one non-relational property, *P*, such that *x* is *P*. If no form of the Principle of the Identity of Indiscernibles is a necessary truth, then the particularity of a particular cannot be constituted by its having a unique nature. Armstrong maintains that the particularity of concrete particulars is constituted by spatio-temporal position. Concrete particulars such as ordinary material objects are distinguished from what Armstrong calls abstract particulars. Not only are all spatial and temporal parts of particulars themselves particulars, but a concrete particular with, for example, both visual and tactual properties will have as parts an abstract particular with only visual properties and an abstract particular with only tactual properties. Abstract particulars with different properties obviously are often in the same place at the same time. Armstrong claims, however, that it is a necessary truth that no particulars with just the same properties can be in the same place at the same time (I, p. 123). Unless it is somehow a necessary truth that any two concrete particulars have some property in common, it does not follow from this, what Armstrong also holds (I, p. 121, II, p. 66), that no two concrete particulars (neither being a spatial or temporal part of the other) can be in the same place at the same time.

I doubt that good reasons can be given for supposing that place-sharing by concrete particulars, or by abstract particulars with the same nature, is logically impossible (see Sanford [1970]). Armstrong seems to have his doubts as well. He does not give any direct arguments; and he adds in a footnote to the italicised assertion

> *Sameness of total position AND sameness of nature do together guarantee sameness of particulars* (I, p. 123). *Perhaps* it can be maintained that, two particulars can both occupy the same total position and have the same nature. But such particulars, to be two, would require some third principle or 'dimension' of individuation of which we know nothing.

I believe that we are familiar enough with a third principle of individuation, a principle of counterfactual differences. We do agree about certain counterfactuals being much better supported than others even while controversies rage about which theories can best account for this agreement. If *a* would have been *F* and *b* might not have been *F* in certain circumstances which could have obtained, then *a* could have been *F* when *b* was not, and *a* and *b* are therefore distinct. Such a counterfactual difference, so far as I can see, does not logically require some actual difference in nature or total position. Given the apparent non-existence in actuality of place-sharing concrete particulars, an example will have to come from science fiction.

When the trigger of an Integrator Ray Gun is pulled, ten feet away from the end of the gun in the direction the gun is aimed a one-inch golden sphere pops into existence. A sphere that begins to exist at a certain place would have begun to exist a bit farther to the left if the gun had been aimed a bit farther to the left. You and I happen to aim our guns at exactly the same point and to pull our triggers at exactly the same time. The resulting object does not behave like a normal one-inch sphere of gold, in which case we would seem to have only a case of causal overdetermination; it is twice as dense as ordinary gold, is less reflective, and so forth. If we entertain the hypothesis that it is a composite object consisting of two golden spheres with the same nature in the same place at the same time, we will have no hope of separating the two spheres. Since they have the same nature and the same position, any mechanical or chemical techniques of separation will affect each of them in just the same way. But even though there is no way now of effecting a separation, we can give sense to the supposition that the two spheres could have been in different places; for you could have aimed your gun three inches to the left without my gun being aimed differently when we pulled the triggers.

Armstrong does not sketch a theory of counterfactuals in this work, but he makes clear his lack of sympathy to any theory that treats possible worlds with ontological respect. Entities such as possible worlds, numbers, timeless propositions, the null-class, unit-classes, and any classes that cannot be identified with the aggregate of their members, have no place in Armstrong's scheme of the world (I, p. 128). According to his world-hypothesis, all first-order particulars (particulars that are not also universals) are spatio-temporal entities existing in a single spatio-temporal system that is completely described in terms of a (completed) physics (I, p. 126). Possibilities and necessities, especially the necessities of logic and mathematics, have to be accounted for somehow on this hypothesis, and Armstrong remarks in several places that his program is not yet completed. Works such as 'Possibilities Without *Possibilia*' and *A Naturalist Philosophy of Mathematics* may be in the making.

While volume II, *A Theory of Universals*, has quite a few negative results, its main purpose is a defence of the theory Armstrong accepts. Part of the accepted theory is the Principle of Conjunctive Universals. Any conjunction of universals is itself a universal so long as it satisfies the Principle of Instantiation. Chapter 14, however, is 'devoted to arguing that there are no disjunctive or negative universals', or so Armstrong says and seems to believe. Without indicating any awareness, while discussing these topics, between asserting a claim of the form *No F's are G's* and rejecting the corresponding claim of the form *All F's are G's*, Armstrong also says 'The theory which will be criticized is that *any* predicate of the form "*P* v *Q*", where "*P*" and "*Q*" are property-predicates, is itself a (disjunctive) property-predicate' (II, p. 20) and 'The hypothesis to be argued *against* may be stated thus. Given that "*P*" is any property-predicate, then, if "$\sim P$" has application, it is a property-predicate also' (II, p. 23). With one exception in each case (an argument from the nature of causation), the arguments that follow do not support the claim that there are no disjunctive or negative universals. For example, Armstrong says that

> disjunctive properties offend against the principle that a genuine property is identical in its different particulars. Suppose *a* has a property *P* but lacks *Q*, while *b* has *Q* but lacks *P*. It seems laughable to conclude from these premisses that *a* and *b* are identical in some respect. (II, p. 20)

The degree of laughability seems to me to depend greatly on just what properties *P* and *Q* are. We can hold that disjunctions of some pairs of property-predicates are themselves property-predicates without holding this for every pair. Similarly, we can hold that there are some pairs of complementary property-predicates without holding that in every complementary pair, one is a property-predicate only if both are. We want not to regard each of a pair of complementary predicates as negative (Sanford [1967]), yet for some pairs there appears to be no reason for regarding one as negative rather than the other. Is opacity the absence of transparency, or transparency the absence of opacity? Is absorbtivity the lack of reflectivity, or the other way round?

Armstrong would like to hold that colours are objective properties, but he does not present a worked-out case for this view. He does mention in passing 'the scientific identification of colours with the sorts of wave-length emitted from the coloured surface' (II, p. 52). The theory scientist actually hold, that the reflectance of a coloured surface is more important in determining its colour than the composition of light it reflects (Land [1977]), appears to support Armstrong's view of colours better than the theory he attributes to scientists.

While always insisting that only science can tell us what properties there are, he provides no illustrations of how the existence of a property is scientifically established. Without definite examples of properties, it is difficult to formulate definite counterexamples to his claims about properties. I shall nevertheless attempt to formulate a counterexample to his Principle of Logical Equivalence,

> If predicates are logically equivalent, but not logically empty, then they apply, where they do apply, in virtue of the very same property or properties (relation or relations). (II, p. 40)

I assume that two predicates do not apply in virtue of the very same property or properties when each predicate applies to something to which the other does not apply. I assume also that a conjunctive predicate applies in virtue of the same properties in virtue of which its conjuncts apply. Now the predicates 'is an equiangular polygon', 'has a 60° angle', 'is an equilateral polygon', and 'has three sides' are each logically independent of the others; yet the conjunction of the first two is logically equivalent to the conjunction of the last two. In general, logically equivalent conjunctions can have logically independent conjuncts. When a property is complex, it should not be assumed that there is just one way of analysing it into simpler properties.

If *P* is a conjunctive property with conjuncts *Q* and *R*, then everything *P* is also *Q* and *R*. Not all components of complex universals are conjuncts. A thing's having a structural property *P* is at least partly constituted by its proper parts having a property distinct from *P*. A thing's having a relational structural property is at least partly constituted by its proper parts having certain relations to each other. Relations are genuine universals. Armstrong thinks it is possible, if not very likely, that every property is a relational structural property (II, pp. 80–4). Some ontologically complex universals are epistemologically simple for us because we can register their presence without being in any way able to analyse them (II, p. 54). Armstrong thinks it possible that every universal is ontologically complex and thus infinitely complex. He also entertains the mind-boggling possibility that an 'infinitely complex property might be a complex of simple constituents, but simples of which there are an infinite number' (II, p. 67). Armstrong's willingness to allow the infinite complexity of universals appears to me to raise a difficulty for his account, to be discussed soon, of resemblances between universals.

Armstrong holds that no genuine relation ever relates a particular to itself (II, pp. 91–3). He realises that if one object is related to another by a relation that is both transitive and symmetrical, as these second-order properties are usually defined, it follows that each object is related to itself; so he suggests that the usual definition of transitivity be emended. The rejection of reflexivity is not so easily dismissed as it might first appear (see Broad [1933], pp. 90–2). The truth of a sentence of the form '*Raa*' is a counterexample only if '*Rxx*' stands for a genuine relation and only if '*Raa*' is not true merely because one part of *a* is related to another part of *a*. It is tempting to argue as follows:

> If *x* and *y*, which each have a size, do not have the same size, then *x* has a size different from the size of *y*. Therefore, if something does not have the same size as itself, but does have a size, it has two different sizes, which is absurd.

One who rejects reflexivity could insist that the premiss here is true only when x and y are distinct, but Armstrong apparently wants to deny that *being the same size as* is a genuine relation. He says that we know *a priori* that a particular is the same size as itself (II, p. 92), and holds that we can never determine *a priori* what universals a particular has. The determination that something is the same size as itself, however, is not purely *a priori*, strictly speaking, since it requires the determination, which is not purely *a priori*, that the thing has some size. At any rate, Armstrong's so-called Irish principle, 'If it can be proved *a priori* that a thing falls under a certain universal, then there is no such universal', seems to me extreme. He does not discuss the possibility of rejecting many *a priori* arguments for the existence of certain universals without rejecting them all.

In part six, 'The Analysis of Resemblance', Armstrong attempts to provide a better account than those offered by any of the theories rejected in volume I. The resemblance of particulars receives a standard Realist account. One particular resembles another if and only if there is a property, P, one has, and a property, Q, the other has, and either $P = Q$ or P resembles Q (II, p. 96). This immediately shifts the problem of accounting for the resemblance of particulars to the problem of accounting for the resemblance of universals. Some resemblances between universals can be accounted for in terms of the properties and relations of particulars. For example, two properties resemble each other in being cloying if any particular that has either property cloys as a causal result of having the property. The standard account of resemblance between particulars applies to some resemblances between universals. For example, two universals resemble each other in being complex if each has the second-order property *being complex*. Armstrong argues that many resemblances between particulars cannot be handled on either of these accounts and also finds deficiencies in R. W. Church's appeal to second-order relations and in W. E. Johnson's theory of determinates and determinables.

Armstrong's own account appeals to the partial identity of universals. Complex universals have parts, and two complex universals which have some, but not all, of their parts in common are partially identical. Armstrong's earlier distinction between epistemological and ontological complexity is important here. Colours, for example, are epistemologically simple although their resemblances are due to ontological complexity. Armstrong does not raise the question whether ontologically simple universals can resemble each other. If they can, their resemblances will obviously have to be accounted for by an appeal to something other than shared components.

Armstrong's one worked-out example of degrees of resemblance between universals concerns lengths that

> exhibit a resemblance-order based upon their intrinsic natures. *Being a foot* resembles *being a yard* more than it resembles *being a mile*. This order is explained by the order which holds among the partial identities. It is simply the transitivity of part to whole. The resemblance may be more or less close, because the partial identity may be more or less close to the limiting case of *complete* identity. (II, p. 123)

Armstrong deals here with only the very easiest sort of case. How, on this account, does one determine what shorter length resembles *being a foot* just as much as *being a foot* resembles *being a yard* or *being a mile* or being any longer length? Does *being a mile* resemble *being a yard* or *being thirteen inches* more than it resembles *being a foot*? Such questions should have objective answers if degrees of resemblance are due to partial identities, since partial identity is an objective matter. Armstrong's earlier speculation that all universals are infinitely complex appears to raise difficulties for his theory of partial identity. If any two universals that share some component thereby share infinitely many components, how can different degrees of resemblance be grounded in different degrees of partial identity? Also, if there are alternative analyses of complex Universals into simpler components, as suggested above, the degree of partial identity between universals will depend on how they are analysed.

Armstrong denies that there are any determinable universals. 'All genuine universals are determinates. There are such predicates as "coloured" or "red", but there is no property, *being coloured* or *being red*' (II, p. 117). This contention is supported by the following argument:

> Suppose that *redness* is a property which all red particulars, whatever their shade of red, have in common. Since properties are universals, this entails that the particulars are *identical* in a certain respect: in respect of their *redness*. Now consider particulars of different shades of red. It is in this very respect of *redness* that they differ. Yet it is impossible that things be identical and different in the very same respect. It is undeniable that different shades of red are different properties. It follows that *redness* is not a property common to all red things. (II, p. 117)

Armstrong's own discussion of length, his other main example in his discussion of resemblances between universals, suggests that there is indeed a way things can be both identical and different in the same respect.

> Select just one member of the class of lengths. All the lesser lengths are structural parts of this length. This length is a structural part of all greater lengths. Assign the number 1 to it. All the lesser and greater lengths may be assigned distinct positive real numbers, numbers which measure the proportion which the lengths bear to length. In this fashion, the class of lengths is united by a single, topic-neutral, formula. (II, p. 128)

Now consider particulars of different lengths. In respect of having proportion to the unit length they are both identical and different. They are identical in each having some proportion to the unit length; they are different in each having a different proportion to the unit length. However, the complexity of redness is to be analysed, we should expect to find analogous respects that can both unite and differentiate different shades of red.

Armstrong apparently assumes that if no universals are determinable, then all universals are determinate. But if any determinable is something in respect of which two particulars can both be the same and different (both red, for example, but different shades of red), this assumption overlooks universals that may be neither determinate nor determinable. 'Absolutely determinate' has two meanings: 'absolutely specific' and 'absolutely precise'. Specificity and precision are independent. If a (more or less specific) universal is not absolutely precise, then there can be a borderline case, something that neither definitely has the universal nor definitely lacks it. Armstrong does not discuss vague universals. His argument that all universals must be determinate is concerned only with specificity. Right after completing his argument, Armstrong characteristically entertains the possibility that his conclusion is false. The world, he reluctantly admits, might be indeterminate in that some particular has a universal but not in virtue of having some absolutely specific universal.

> Suppose that the length of a certain particular at a certain time is said to be indeterminate over a certain range. It has L_1 or L_2 . . . But this is not because it has, say, L_1, while lacking any other of the disjuncts. Nor is it because the particular has each one of the disjuncts. If we allow this as a possibility, then we require 'disjunctive' properties which do not obey the truth-table for classical disjunction. (II, 119)

In entertaining the possibility that there are absolutely specific lengths that no particular has, Armstrong appears ready to abandon his Principle of Instantiation. He could instead abandon the assumption that every universal that is not absolutely specific is equivalent to a disjunction (classical or otherwise) of more specific universals. If, for some number n, a man is n (rather than $n-1$ or $n+1$) millimetres tall, and there is no number m such that he is m (rather than $m-1$ or $m+1$) Angstroms tall, then why should his height, to the nearest millimetre, be regarded as a huge disjunction of heights to the nearest Angstrom? On Armstrong's own theory of complexity, it is difficult to see how all universals can be absolutely specific. If a conjunctive property is more specific than its conjuncts, as it apparently must be, then its conjuncts are less specific than it. A complex universal will generally be more specific than its components, so its components will be less specific than what they compose.

The last topic of the book is higher-order universals. Universals can have both properties and relations. Armstrong allows that second-order relations might have third-order properties, although he doubts that there are any other universals above the second order. All higher-order universals are formal or topic-neutral. An extremely important second-order relation between universals is *nomic probabilification*, which admits of degrees. In its lowest degree, it is nomic exclusion. In its highest degree, it is nomic necessitation. All deterministic causal sequences are cases of nomic necessitation, and causal sequences in accord with probabilistic laws involve lower degrees of nomic probabilification.

Armstrong's theory of universals and his theory of causation are of a piece. Earlier in volume II Armstrong suggests that

> (a) the active and passive powers of particulars are determined by their properties. (b) Every property bestows some active and/or passive power upon the particulars of which it is a property. (c) A property bestows the very same power upon any particular of which it is a property. (d) Each different property bestows a different power upon the particulars of which it is a property. (pp. 43–4)

A property bestows a power because of its second-order relation to another universal. These suggestions combined with Armstrong's theory of causation thus imply that all universals are related to other universals.

Nomic necessity entails constant conjunction, although constant conjunction does not entail nomic necessity (II, p. 153). In admitting the reality of the second-order relation of necessitation, Armstrong rejects Humean analyses of causation, but he wishes to preserve certain Humean insights. He claims that nomic necessitation is never logical necessitation. I have suggested elsewhere that kinematics provides examples of necessity that are both causal and *a priori* intelligible ([1975], [1978]). The supposition that all causal necessitation is ultimately *a priori* intelligible may be too good to be true, but it is not refuted by our inability to figure out *a priori* the causal powers of water, bread, and so forth merely from their sensory qualities. Armstrong insists that we are often aware of complex universals without being aware of their complexity. If this is so, it is no wonder that relations between universals we have yet to analyse can be discovered only empirically. He is confident that even when complex universals are fully analysed, 'there will not even appear to be logical links between nomically linked universals' (II, p. 152) but offers little argument for this bit of modern orthodoxy. At the very end, after a brief discussion of our direct awareness of nomic necessitation, he appeals to the principle that there are no logically necessary connections between distinct existences (II, p. 166). This is what he calls a 'fundamental principle of Empiricism' in [1961], p. 71. The difficulty of understanding 'distinct existences' so that this principle is neither vacuous nor obviously false appears to be increased by accepting an account of resemblance between universals in terms of partial identity. Armstrong's credentials as a good Empiricist do not require his continued homage to this so-called fundamental principle.

Universals and Scientific Realism contains so many interesting and provocative arguments and suggestions that I have allowed my own preoccupations and heterodoxies to guide my selection. Other reviewers will single out passages I have neglected. Although many of my comments have been critical, I want not to convey the impression that this is a largely unpersuasive work. I feel about Armstrong's theory of universals (as I do about his theories of perception, mind, and knowledge) that despite its difficulties it is closer than its competitors to being adequate. Anyone with an interest in the topic who reads this book will learn, will be persuaded of something, and will be stimulated by seeing a productive philosophic mind in action.

At the end of each volume there is a glossary of terms and principles, a list of works cited, and a good index. The glossary distinguishes rejected from accepted principles, so one can get a good idea of Armstrong's views by reading through it. The length of the list of works cited (over 5 pages) attests to his scholarship on the topic. The material at the end of volume I is all repeated at the end of volume II, and the table of contents at the beginning of volume II is repeated from volume I. Each volume briefly summarises the argument of the other. When these summaries and the repeated material are omitted, the whole work runs to xvi+315 pages. This is not long enough to justify its division into separately bound volumes. Few will want to read one without reading both. In these inflationary times, with book prices going up faster than the prices of many goods, book reviews often end with complaints about high costs. This review ends with such a complaint in the hope of discouraging the expensive practice of stretching one book into two.

DAVID H. SANFORD
Duke University

REFERENCES

ARMSTRONG, D. M. [1961]: *Perception and the Physical World*. London: Routledge and Kegan Paul.
BROAD, C. D. [1933]: *Examination of McTaggart's Philosophy*. vol. I. Cambridge: Cambridge University Press.
EVANS, G. [1978]: 'Can There be Vague Objects?', *Analysis*, **38**, p. 208.
KRIPKE, S. [1971]: 'Identity and Necessity', in M. K. Munitz (*ed.*): *Identity and Individuation*, pp. 135–64. New York: New York University Press. Reprinted in S. P. Schwartz (*ed.*): *Naming, Necessity, and Natural Kinds*, pp. 66–101. Ithaca: Cornell University Press.
LAND, E. H. [1977]: 'The Retinex Theory of Color Vision', *Scientific American*, **237** (6) (December), pp. 108–28.
SANFORD, D. H. [1967]: 'Negative Terms', *Analysis*, **27**, pp. 201–5.
SANFORD, D. H. [1970]: 'Locke, Leibniz, and Wiggins on Being in the Same Place at the Same Time, *Philosophical Review*, **79**, pp. 75–82.
SANFORD, D. H. [1975]: 'Causal Necessity and Logical Necessity', *Philosophical Studies*, **28**, pp. 103–12.
SANFORD, D. H. [1978]: 'Causal Necessity and Logical Necessity (*Reply to Brand and Swain*)', *Philosophical Studies*, **33**, pp. 185–94.

Brit. J. Phil. Sci. 31 (1980), 81–108 *Printed in Great Britain*

Reviews

CARROLL, LEWIS [1977]: *Symbolic Logic*, edited by William Warren Bartley, III. New York: Clarkson N. Potter, Inc. $14.95. Pp. xxv + 496.

Professor Bartley has prepared a handsome hard-cover edition of Lewis Carroll's writings on symbolic logic, illustrated with well-reproduced photographs of Carroll, Carroll's sister, and his long-time colleague Thomas Vere Bayne. It also contains a variety of cartoon illustrations, some diagrams, and facsimile reproductions of some of Carroll's letters, some pages from his logic notebooks, and some early galley proofs of the not heretofore published *Symbolic Logic* Part Two.

Readers acquainted with the Dover Fourth Edition (1958) of Carroll's *Symbolic Logic* Part One will recall that it is divided into eight 'Books' and an Appendix. The new Part Two, unfinished at the time of Carroll's death in 1898, has been assembled by Bartley from scattered early galley proofs and manuscripts. It also is divided into 'Books', numbered IX through XXII by Carroll himself, who must have planned to have both Parts ultimately published in a single volume. (There is some reference to a projected Part Three, but nothing identifiable as part of it has survived.)

Part One begins like most traditional logic texts: Book I deals with (class) *terms*, including definition by Genus and Differentia, and Book II with *propositions*, whose normal forms are analysed into the familiar Sign of Quantity, Subject, Copula, and Predicate. Carroll treated 'All S is P' as a 'Double Proposition', asserting both that no S is other than P and that some S is P. Book III introduces Carroll's ('biliteral') diagrams for single propositions: these are rectilinear in contrast to the more familiar curvilinear Venn Diagrams, and differ from the latter also in having the 'Universe of Discourse' closed rather than open. Book IV advances to 'triliteral' diagrams which—like three-circle Venn Diagrams—permit representation of three propositions containing three distinct terms. Finally we arrive at *arguments* in Book V, which deals with (categorical) syllogisms, whose validity is tested using 'triliteral' diagrams. In Book VI an algebraic representation of propositions is introduced by way of juxtaposing lower case latin letters that represent terms, together with accent marks for term complements, and subscripts 1 and 0 to indicate that the classes or class products represented are respectively nonempty ('existing') or empty. Two propositional connectives are introduced also: a dagger for conjunction and a (reversed) paragraph symbol for implication or entailment. This symbolic apparatus permits Carroll to develop three general formulae representing valid syllogistic inference patterns. Similarly, three 'Forms of Fallacies' are identified. In Book VII the sorites is considered. Carroll's problem here is not to determine the validity or invalidity of a given sorites, but rather: given three or more propositions as premisses, '. . . to ascertain what Conclusion, if any, is consequent from them.' To solve this problem a method of cancelling terms by underscoring is described. Book VIII consists of 'Examples, Answers, and

Solutions': in it nearly 400 problems are posed and their answers or solutions given, using diagrams, rules, and the method of cancelling terms in formulas. This edition contains more solutions than previous editions did, supplied by the editor partly from Carroll's own manuscript annotations on a copy of *Symbolic Logic* Part One and partly from a galley sheet preserved at Oxford University.

Part Two, labelled 'Advanced', is not altogether new, as the editor explains on pages 11–12 of his Introduction. Book IX consists of only one page, taken from section 10 of the Appendix to Part One from the Dover Fourth Edition, and Book X consists of the bulk of the remainder of that Appendix. Book IX indicates briefly what Carroll intended to take up in Parts Two and Three. Book X contains Carroll's views on the existential import of propositions, his reflections on the Copula, a brief presentation of Euler's and Venn's diagrams, and the way in which his own diagrams can be extended to deal with arguments containing as many as ten terms. Book XI develops five patterns of valid inference, some syllogistic and some involving propositions of the form 'Not all *a* is *b*'. Because Carroll interprets 'All *a* is *b*' as a conjunction, its negation is a disjunction, and the symbol '§' is introduced to express 'or'. Four fallacies are also identified and given names. Seven 'Logical Charts' are presented and (partially) explained by the editor, who remarks that they provide 'a handy ready reckoner' for inferences involving propositions of the 'Not all *a* is *b*' form. Bartley confesses that he has not been able to figure out all of the seventh chart, and on page 273 expresses the hope that 'Perhaps one of the readers of this book will be able to decipher it.' Book XII presents Carroll's 'Method of Trees', which utilises the *reductio ad absurdum* principle to establish validity for syllogisms and sorites that have universal premisses and conclusions. It has some resemblance, as the editor points out, to Beth's Semantic Tableaux, though it is more cumbersome and very much more limited. Book XIII contains problems to be solved by the methods of Part Two: some from the Appendix to Part One, some from Carroll's privately printed papers on logic, and some from 'materials designated by Carroll for Part II but not sorted out and classified into books' (p. 12). It also contains hints on how to solve the problems, some by Carroll and some by the editor, as well as some complete solutions provided by the latter. In it are also some letters from Carroll to his sister, Louisa Dodgson, and to Professor John Cook Wilson, dealing with solutions to the problems presented. Book XIV contains nearly seventy more problems, including one mammoth sorites of fifty premisses, and several others nearly as long. Then there is a gap followed by Book XXI with the title 'Logical Puzzles'. These include the liar paradox and its crocodile variant, Achilles and the Tortoise, Carroll's own 'What the Tortoise Said to Achilles' (reprinted from *Mind* 1895), and his Barber-Shop Paradox (much the same as the version printed in *Mind* 1894). There are also Carroll's solutions to the puzzles, which do not seem to break any new ground. Three Appendices are attached to this Book. In Appendix A Bartley discusses various proposed 'solutions' to the Barber-Shop Paradox; Appendix B offers a 'variorum' treatment of Carroll's eight versions of it; and in Appendix C Bartley gives a very reasonable discussion of 'What the Tortoise Said to Achilles', in which he disputes the view of Gilbert Ryle and Stephen Toulmin that statements of 'laws of nature' express 'inference licenses' rather than propositions. Finally, Book XXII, of which only Chapter I survives, is titled 'Solutions of Problems Set by Other Writers'. The single chapter reproduces twenty-five problems culled from the writings of Boole, DeMorgan, Jevons, J. N. Keynes, and others. Some of them are accompanied by comments, some by solutions.

Also present in this volume (p. 49) is Carroll's denial of 'a silly story, which has been going the rounds of the papers, about my having presented certain books to Her Majesty the Queen'. (A version of that story can be found in Alexander Woolcott's Introduction (p. 4) to the Modern Library Giant's (misnamed) *Complete Works of Lewis Carroll.*) But absent from the present volume is Carroll's *The Game of Logic*, a ninety-six page book that was bound together with the Dover Fourth Edition of *Symbolic Logic.*

Although Bartley tells us that 'Carroll's examples and exercises manifested genius' (p. 29), I suspect that different readers will have different levels of appreciation of the many illustrations and problems in this volume. One adult reader of my acquaintance chortled uncontrollably over syllogisms and sorites dealing with illogical babies who cannot manage crocodiles (p. 160), contemptible fish that cannot dance a minuet (p. 173), and guinea-pigs hopelessly ignorant of music (p. 165). Another, exposed to this unremitting whimsy, was moved to quote Dorothy Parker's review of a book by A. A. Milne: 'Tonstant weader fwowed up.' Still others, despite the Note from the Editor and Publisher on page vii on the 'historical setting' of Carroll's examples, may be put off by being told that 'No niggers are white' (p. 149), and by the varied remarks about Jews: that they are sarcastic, either hunchbacked or misers; obsequious unless very young, squinting, dishonest, look like goats, have beards a yard long and hooked noses, and say "shpoonj" (pp. 403, 405, 416, 393, 156, 262, 397, 150, 163, 161).

In addition to the Appendices mentioned, the editor, Professor Bartley, has provided copious notes and useful comments throughout the text, prepared a bibliography (pp. 39–42), and wrote a lengthy Introduction (pp. 3–36). Part I of the Editor's Introduction tells briefly about Carroll himself and his writings on logic. Part II reports how Bartley became interested in finding the lost Part Two of *Symbolic Logic*, his lengthy search for its far flung parts, and the plan followed in assembling them into their present arrangement.

In Part III of the Introduction Professor Bartley discusses the history of logic. I think he is right in emphasizing the radical discontinuity between the 'second period' (Boolean or algebraic logic) and the contemporary period of mathematical logic. As he points out, the typical problem posed by textbook writers of the second period consists of a number of conditions (premises) from which the student is required to draw his own conclusion by algebraic methods, whereas the typical problem in a contemporary textbook presents a complete argument (premises and conclusion all stated explicitly) that the student is required to prove valid or to prove invalid by methods other than the algebraic. And as Bartley remarks, more advanced textbooks and treatises today investigate the foundations of mathematics and the general characteristics of formal systems. I think his case could have been made even stronger by contrasting the *aims* of the two periods. The aim of Boole and his followers was to devise symbolic techniques that permit more effective inferences in the logic of terms, of which the categorical syllogism is the simplest example. But starting with Frege, the aim of mathematical logicians was to devise symbolic techniques that can be used to analyse and validate derivations within mathematics itself. This latter aim required both quantification theory and the logic of relations, neither of which is required for solving the typical problems considered by Boole, Venn, Jevons, J. N. Keynes, and Lewis Carroll.

On the other hand, I think that Bartley exaggerates the usefulness of algebraic logic when he writes that 'Logic as presented by Carroll is . . . a kind of instructional aid, of obvious pedagogical utility, for detectives' (p. 24). Where a detective would *get* fifty relevant premisses as in Carroll's problem on pages 386–8 is itself a puzzle. I am inclined to agree with Passmore's judgment that the algebraic logicians' illustrations and exercises '. . . were for the most part distinctly artificial problems, not likely to arise in any actual inquiry' (*A Hundred Years of Philosophy*, p. 148).

Bartley makes a somewhat surprising claim about the relative contributions that algebraic and contemporary logical training can make to the general philosophic enterprise. On page 25 we are told that

> most philosophical problems require, for their solution, the kinds of deductive and analytical skills for which Carroll, Jevons, Venn, and their contemporaries invented their puzzles, and do *not* require the metamathematical theory and techniques of contemporary logic.

To support this claim Bartley states that

> Although contemporary philosophers are given a specialized education in mathematical logic, their ordinary work in philosophy is littered with elementary logical mistakes. Non sequiturs abound (p. 26).

As evidence Bartley adduces two bad arguments published by two contemporary philosophers, neither of whom would proclaim himself to be a special expert in mathematical logic. Mistakes are all too frequently made, and they are not confined to contemporary philosophers. (Bartley remarks on a somewhat similar boner made by Carroll himself (p. 449), but in his general discussion he does not mention this evidence that is contrary to his claim.) Better evidence for Bartley's claim would have been a number of philosophical problems actually solved by Carroll, Jevons, or Venn using their algebraic method. But no such evidence was adduced. To the best of my knowledge no such evidence exists.

Bartley propounds an interesting theory about twentieth century philosophy. The success of Russell in resolving the logical (and semantical) paradoxes by introducing type-theoretic canons of meaningfulness made 'many twentieth-century philosophers' suppose that such techniques 'could be extended *beyond* formal logic to the traditional problems of philosophy . . . [which thus] . . . could be made to disappear' (p. 26). But this supposition is flatly wrong, Bartley argues, 'For the self-reference that is to be found in the logical antinomies is *simply absent* from most traditional philosophical problems' (p. 26). No doubt there were *some* philosophers who made this mistake. But most philosophers who tried to dissolve traditional problems by using a criterion of meaningfulness appealed to a 'principle of verifiability' rather than any notion of self-reference. I think there is some confusion here—not entirely or exclusively Bartley's, of course. Indeed, when one looks at the attention paid to traditional philosophical problems by those whose credentials in mathematical logic are beyond question, the situation is otherwise than Bartley suggests. Russell's 1903 discussion of Zeno and continuity (*Principles of Mathematics*, pp. 346–58) attempts to solve rather than to dissolve the problem of change; and his 1905 essay 'On Denoting' uses the methods of contemporary logic to solve Parmenides's problem of how we can talk about what doesn't exist. Quine's brilliant article 'Identity, Ostension, and Hypostasis" in *From a Logical Point of View* includes his solution of Heracleitus's problem of bathing in the same river twice. And Gödel has taken the 'Problem of Universals' very seriously, even coming down boldly on the side of Platonism. It is certainly understandable that an editor should stress the importance of the material he edits. But in my judgement Professor Bartley's claims here have been exaggerated, and his arguments unconvincing.

I am puzzled that Bartley can say so casually that Carroll's 'Oxford contemporaries were . . . in part flirting with psychologistic logic under the influence of F. H. Bradley" (p. 30). It is true that Bradley's notion of an "ideal content, may suggest psychologism, but Bradley writes at the beginning of his important and influential *Principles of Logic* that in England '. . . we have lived too long in the psychological attitude' (p. 2). Most commentators on Bradley have emphasised his implacable opposition to the psychologism of the British empiricists, especially Mill. H. B. Acton in the Crowell Collier and Macmillan *Encyclopedia of Philosophy* wrote that 'Bradley set out to refute . . . psychologism' (vol 1, p. 360). Gilbert Ryle in *The Revolution in Philosophy* wrote that 'Frege and Bradley in different ways and with different emphases distinguished sharply between psychology on the one side and philosophy and logic on the other . . .' (p. 6). In that same book Richard A. Wollheim discusses in detail 'Bradeley's severance of logic or philosophy from psychology . . .'(p. 13) and in much greater detail in his book on Bradley (see especially Chapter 1, 'Logic and the Rejection of the "Psychological Attitude"').

Much of what I have expressed disagreement with Bartley over is a matter of judgment and interpretation. But there is one point that I can correct him on with complete assurance. On page 447 he refers to 'causal implication as developed by Burks and Copi . . .', I can say with confidence in my knowledge of the situation that all the credit for developing the logic of causal implication belongs to Professor Arthur W. Burks, and none at all to

IRVING M. COPI
University of Hawaii at Manoa

MANIER, E. [1977]: *The Young Darwin and His Cultural Circle. A study of influences which helped shape the language and logic of the first drafts of the theory of natural selection.* Dordrecht: D. Reidel. Hardback: $24.50. Paperback: $11.95. Pp. xii + 242.

To clarify and evaluate what Manier has attempted and achieved in this book, it is necessary to introduce very briefly the chronology and documents essential to any interpretation of Darwin's thinking in the five wonder-years (1837–1842) in Cambridge, London and Kent, that followed the five wander-years (1831–1836) of the voyage of H.M.S. *Beagle*. When Darwin came off the *Beagle* in October 1836 his immediate duties were to write up a narrative of his travels and observations, and to arrange for various specialists to study his collections of material, the animal and bird skins and fossil bones especially. A draft of the narrative, published in 1839 as his *Journal of Researches*, was complete by June 1837. The zoological and paleontological work was going ahead well, too, by then. So July 1837 was an appropriate time to open notebooks devoted explicitly to his speculations concerning such general theoretical problems as the origin of species. These speculations had been increasingly on his mind all through that year, but he had made only brief notes and nowhere worked out in detail his thoughts on what he called 'the laws of life'. The four notebooks that he eventually filled with these reflections run from July 1837 well into 1839. From their opening he was exploring the causes and effects of the transmutation of species. It is, however, only late in 1838 that he seems to have had any notion of new species arising by means of natural selection. The exact moment when this conjecture first occurred to him is uncertain. It may have come immediately upon reading Malthus late in September that year; perhaps it came only after several more weeks. In any case natural selection was definitely established in Darwin's mind by the end of 1838; although, tantalisingly, there are only short scrappy formulations of the theory until 1842, when Darwin wrote out his *Sketch*, the comprehensive exposition, of which the *Essay* of 1844 and *Origin of Species* of 1859 are longer, more elaborate versions.

We lack, then, extensive direct documentation either of Darwin's first formulations of natural selection or of his earliest thoughts on how to argue to and from the theory in making the best case for its evidential credibility and explanatory virtue. What we do have, however, is another pair of notebooks—the first was opened in July, 1838—devoted to what Darwin called metaphysics (that is, primarily, metaphysics in the Scottish philosophers' sense of psychology very broadly understood) and morals. These notebooks concern the special challenges raised by an animal ancestry for our own species, and Darwin takes up in them such topics as the origin of language, of aesthetic and ethical values, of instincts for sociability and for religious sentiment.[1]

Now these six notebooks, when taken all together with the other contemporary manuscript material, and the printed works relating to South American natural history and geology written in these five years between the end of the voyage and the writing of the *Sketch*, present a picture in striking and refreshing contrast with the one associated with Darwin later in life and the one fostered notably by his own *Autobiography*. Instead of an ideologically timid and philosophically naive naturalist who often left it up to others to draw the wider implications from his books on the *Origin of Species* and *Descent of Man*, we find, at the time he was first developing the ideas those works were later to expound, another sort of man and another sort of mind altogether. With an intellectual zest and toughness a hagiographer might be tempted to compare with the physical energy and courage he had shown on his travels, he worked seriously and strenuously if not always deftly and decisively to relate his many new ideas on species in general and on man in particular to the main traditions of thought on providence, evil, freewill, necessity, causation, innate ideas, language, ideals of explanation and evidence and the sources of morality.

The young Darwin revealed in these early documents obviously confronts us with major historiographical challenges. There are dire difficulties of the chicken and egg sort that often involve in turn intricate puzzles over paleography and chronology. These are difficulties that some people may find it hard to face fairly and squarely. Anyone committed generally to the view that biological theorising is always ultimately theorising about man and society will obviously want to see Darwin's solutions to problems in biogeography or embryology as somehow derivative from more fundamental concerns with theodicy or political economy. Equally, anyone upholding the view that physical science makes genuine progress only when disentangled from distracting controversy about reality, knowledge and the good will tend to write off many of Darwin's speculative reflections as fruitless, peripheral accompaniments to secure conclusions already established independently upon their own ground.

---

[1] These metaphysical notebooks are now published in full in H. E. Gruber and P. H. Barrett [1974].

A main virtue of Manier's book is that it keeps these larger interpretive issues well to the fore throughout. A principal weakness is that the need for careful, incisive explication of these issues, in close relation to detailed, precise exegesis of the relevant documentation is never properly met. As a result his own chosen ambitions in this book are far from realised. He does, however, succeed, while failing in his grand design, in making a number of valuable contributions of a more limited kind.

Manier's primary ambition is simple enough to identify. He wants to explain why Darwin formulated and argued for his theory of natural selection as he did in the *Sketch* of 1842. His solution is also clear enough in outline. In the year before and the year after first arriving at the theory, probably late in 1838, Darwin read and took seriously some dozen and a half authors writing on literary, religious, political, philosophical and ethical subjects. From these authors—the best known include Lyell, Lamarck, Mackintosh, Whewell, Malthus, Hume, Comte, Herschel, Stewart, Paley, Wordsworth and Ferrier—Darwin derived not merely numerous specific suggestions for his reflection on plant, animal and human origins, but also, Manier proposes, a general methodology and a general theory of metaphor.

On the methodology side Manier often has interesting and sensible suggestions to make. He has, however, been unlucky in missing, sometimes only narrowly, several of the points essential to success in this area. He has realised quite rightly that the division of the *Sketch* into two parts reveals much about Darwin's implicit method. But he has misread the relationship between the two parts and so misconstrued that method. Again he has sensed that Herschel's and Lyell's notions of *verae causae* were important for Darwin, but again he has misunderstood how Darwin formulated his argumentation for natural selection so as to meet the requirements of the *vera causa* principle.[1]

On the metaphor side Manier works hard to distinguish various senses of, especially, the words 'struggle' and 'selection', and future discussion of Darwin's uses of these terms will have to take Manier's distinctions into account. He fails, however, in his attempt to identify a coherent theory of metaphorical discourse behind Darwin's argumentative practice. So, inevitably he fails too in his attempt to find the sources for such a theory in authors like Smart, Paley, Stewart and Wordsworth whose theories or practices regarding metaphor are supposed to have set decisive precedents for Darwin's own.

To some extent Manier's failure, in this general quest for a definite Darwinian view of metaphor, is compensated by the many new and informative reports he has to make on Darwin's voracious reading and on the reflections arising from it. In the case of Hensleigh Wedgwood with whom Darwin discussed materialism, for example, we get very important notes published here that give a wonderfully direct access to Darwin's thinking on mind, matter and bodily organisation. Likewise, in the cases, especially, of John Fleming, James Mackintosh, John Barclay, Benjamin Smart and William Abercrombie, Manier is often an instructive guide as he documents their views and Darwin's responses, negative, positive and mixed as they were. It should be stressed though that Manier's larger ambitions often lead him to go far beyond what the documents will support, in ways that could seriously mislead the unwary. There is, for example, no direct evidence that Darwin derived anything from Smart's book on metaphysics except a confirmation and extension of views about the origin of language that Darwin already inclined to himself before reading Smart. Manier, however, would make Smart also a vital source for Darwin's views on a whole raft of larger issues about the nature of metaphysical theories and the uses of metaphor in scientific explanation.

---

[1] I have discussed some of these points briefly in my [1977].

At other times, we are left wondering why Manier has devoted so little time to certain documents. He does discuss some of those brief and often oblique accounts of natural selection that we find in Darwin's notebook entries late in 1838 and in 1839. There is nowhere, however, a systematic survey of them, much less a careful exegesis of their argumentation. Nor do we get any serious discussion of Darwin's notions from 1838 on concerning artificial selection and its place in the making of animal breeds and plant varieties. It is hard to see why Manier does not feel the need to explore such things.

On one major point Manier is surely right. As David Hull and others have often insisted, the theory of natural selection does not seem to fit easily and obviously with the strategies of explanation and confirmation dominant in much of the physical science of Darwin's day, nor perhaps in our own. Manier would suggest that we try to resolve this philosophical problem partly through historical research into Darwin's own understanding of his theorising and its relations with contemporary theories in physics, economics and psychology. That suggestion is entirely correct and most salutary. Any philosophical understanding of natural selection as a scientific theory must start with such historical research. Where Manier should not be followed is in his hollow, gimmicky hypostatisation into a 'cultural circle' of the people Darwin read (they have nothing else of interest in common) on elevated topics like language, as contrasted with lowly subjects like animal hybrids or corals; in his refusal to look as closely at the empirical problems Darwin's theory was designed to solve as at the metaphysical problems that it was seen to raise, and in his willingness to substitute literary waffle and cliché, about Darwin's supposedly Wordsworthian worries about the meaning of life, for closer study of Darwin's earliest thoughts on natural selection and of his views on their relation to the contemporary philosophies of science.

Altogether, then, I would suggest that Manier's study makes many sensitive and sophisticated contributions to the history and philosophy of Darwinian science, but I would want to suggest that it be read with a healthy scepticism about its principal pretensions.

M. J. S. HODGE
University of Leeds

REFERENCES

GRUBER, H. E. and P. H. BARRETT [1974]: *Darwin on Man* New York.

HODGE, M. J. S. [1977]: 'The Structure and Strategy of Darwin's "Long Argument"', *British Journal for the History of Science*, 10, pp. 237–46.

MATTESSICH, RICHARD [1978]: *Instrumental Reasoning and Systems Methodology.* Dordrecht and Boston: D. Reidel. US $35. Pp. xxii+396.

A convincing case can be made that most of the exciting developments in the social sciences are taking place on what might be characterised as the extremely applied frontiers. A prime example is the newly emerging field of evaluation research, the gestation of which was hastened in the U.S.A. when Congress mandated that all social programmes (covering the broad health, education and welfare areas) should allocate a portion of their massive budgets to the assessment of their own effectiveness. Social scientists have thereby been forced to grapple with a host of fascinating methodological (and other) problems, and psychologists, sociologists, economists, political scientists, statisticians, educational researchers, and even philosophers have rolled up their sleeves and pitched-in. Names like Lee J. Cronbach, Donald Campbell and Michael Scriven will doubtless quicken the pulses of readers of *B.J.P.S.* More familiar, and only somewhat less exciting fields, are policy analysis and the related 'decision sciences', and the various branches of economic analysis.

So it is with some excitement that one approaches Mattessich's large volume, for it bears the promising sub-title 'An Epistemology of the Applied and Social Sciences'. Sure enough, the tantalising flavour also emerges in the introductory chapter, where there is a prominent section 'Towards a Philosophy of Applied Science'. Unfortunately the honeymoon period is short-lived, for the 'Towards' turns out to lead to a somewhat modified—but hardly strengthened—version of the holistic systems approach. Could it be possible for such a hoary old chestnut to be the solution to *anything*, let alone the very real issues facing workers in the sorts of fields mentioned earlier?

Further exploration of the volume does not inspire one to give an affirmative answer, but certainly there are a lot of surprises in the course of the journey. The reader has to traverse a lengthy discussion of the growth of the field of logic, complete with Aristotle, Boole, Russell and Whitehead, Gödel's Proof, Tarski and Carnap. Then one journeys through the realms of predicate, multi-valued, modal and deontic logic; this is followed by a fifty-page encounter with controversies in inductive logic, where one meets with Hume, Mill, Von Wright, Carnap, Russell and Goodman. The by-now somewhat bleary-eyed explorer continues through decision theory, aspects of economic theory, and recent controversies in philosophy of science. Finally, nearing the end of the road, one encounters again von Bertalanffy and his General Systems Theory (*GST*), and *via* a discussion of the big-bang theory in cosmology one reaches the brief exposition of 'five ontological assumptions of systems methodology.' The purpose of these several hundred pages is not perfectly clear; Mattessich merely claims in his introductory remarks that 'the applied scientist must be acquainted with the major issues of up-to-date scientific and philosophic thinking' (p. 12).

But there is something else going on. Everything is grist for the mill of those who use a systems approach; after all, systems—like the poor—are everywhere. Thus, Mattessich treats logic and theories in science as 'systems', and the suggestion appears to be that many of the breakthroughs in these fields (which he outlines in his long expository sections) actually were triumphs for 'systems thinking'. Bertrand Russell, in grappling with his notorious paradox and developing the theory of types, was an incipient-systems-man, and so were many of the other luminaries discussed in the book.

This is too good, of course, and one becomes rather sceptical. Systems of logic, and scientific theories or paradigms, are evidently examples of 'systems'; but so is the primordial lump that exploded during the cosmological 'big-bang'; and also mitochondria and cells; societies and the institutions within them must also be cited as clear cases of 'systems'. The effect is merely to rob a perfectly good word of clear meaning.

The interesting thing is that there are passages in the book which indicate that Mattessich is aware of the problem. He rejects von Bertalanffy's *GST* which aims at 'unification and peculiar reconstruction of most, if not all, sciences' (p. 277); von Bertalanffy was seeking theory which would apply to *any* system (that is, to any group of elements in interaction). In Mattessich's view, this 'goes far beyond' the limits of a reasonable systems approach, and it is *radically* holistic (p. 277). But in definining a system as at least two elements in inter-relation and in insisting that any system is embedded in an environment with which it interacts (pp. 28–37), and in discussing examples from logic, philosophy of science, physical theory, and in citing cells, societies and physical entities as all being systems, it is not clear in just what respects Mattessich's position actually remedies the defects in von Bertalanffy's *GST*. There can only be two candidates here. In the first place, Mattessich recognises that his 'systems approach' is not a radical departure from the traditional approaches found in various sciences, whereas supporters of *GST* have often thought they were doing something completely new (and have often made nasty asides about 'simple minded Galileanism' and so forth). Second, Mattessich expounds a method for 'neutralising' a system, which in essence consists in drawing the boundaries of a system so that certain elements (such as value-judgments) are excluded from it and become located in the environment (or 'super system'). Of course, scientists have always been using this procedure; what is new is to find this being advocated by a writer who professes to take relations seriously (even, perhaps, in an Hegelian 'internal relations' sense). One cannot have one's cake and eat it too; either relations are vital, and therefore all things that are interrelated must be included in the system under study (which is the thesis that supporters of *GST* profess to accept above all else), or otherwise relationships can, on occasion, be severed in which case the systems approach becomes the traditional (and much abused—by systems folk) 'atomistic' or 'analytic' method.

In summary, on laying down Mattessich's volume one cannot help reflecting that he has not overcome a set of problems he himself clearly identifies: 'The systems approach has still to shake off a faddish flavour, and has hardly yet matured into a generally accepted methodological paradigm' (p. 276). In fact, it appears that Mattessich has not even given grounds for suspecting that the approach—whatever it is—will *ever* shake off these defects.

D. C. PHILLIPS
Stanford University

PUTNAM, HILARY [1978]: *Meaning and the Moral Sciences.* International Library of Philosophy and Scientific Method. London: Routledge and Kegan Paul. £4.95. Pp. 145.

This book contains six John Locke lectures delivered at Oxford University in 1976 (most of Lecture II and the first part of Lecture III having been published as 'What is "Realism"?' in the *Proceedings of the Aristotelian Society* 1975–6). There are also three other essays: 'Literature, Science and Reflection' which was written before the Lectures, and 'Reference and Understanding' and 'Realism and Reason' just after. Despite the title, the book is mainly about meaning (especially truth, reference and understanding) and the conflict between realism and idealism, and only secondarily about the moral sciences. The themes of the book are of current interest in philosophy and Putnam brings a number of important considerations to bear on them. However, the book is not easy reading because most of the pieces retain the style of lectures. The final piece 'Realism and Reason', when delivered as an APA address, may have been assisted then by an explanation of the three diagrams (two the same) that are reproduced in the book but which are nowhere explained. Nor is it always clear what are the arguments that support the positions that Putnam wishes to adopt. However, no prospective reader should be deterred from studying this brief but wide-ranging book.

The first lecture begins with a point raised by Hartry Field[1] concerning Tarski's theory of truth. Briefly, this is that the theory depends on a notion of reference which is defined for English as a list of the following form:

> $(t)(x)[t$ refers to x *iff* either t is 'Plato' and x is Plato, or t is 'Aristotle' and x is Aristotle, or . . .]

But Field argues that to provide such a list is to fall short of higher standards of what constitutes a theory of the notion of reference. What form would a theory of this word-world relationship take? Field's answer is that we should adopt a physicalist approach and pursue a theory of the form

$$(t)(x)[t \text{ refers to } x \textit{ iff } D(t,x)]$$

where the extension of '$D(t,x)$' is determined by facts of physics, neurophysiology, sociology, *etc.* Such a physicalist, or causal-explanatory theory, would, like any other theory of reference, have to rule out as trivial lists of the above kind as determining the extension of '$D(t,x)$'.[2] Clearly Field envisages a reductive explanation of reference along physicalist lines in much the same way as the physicalist approach has reduced the concepts of temperature of phenomenological thermodynamics or valency of chemistry.

Putnam does not argue directly against Field's account of what form a theory of reference should take. He cites some arguments from an unpublished paper of Stephen Leeds, but we cannot gleen too much of Leeds' reply from what Putnam tells us. The gist seems to be that reference is not a causal-explanatory notion, such as valence, which can be given a reductive account along physicalist lines. He contends that 'we can give a "transcendental argument" for Tarski's procedure by appealing to a *purpose* for having notions like *truth* and *reference* which is not at all parallel to the purpose for which we have notions like *valence*' (p. 16).[1] What might this be like? A suggestion can be found elsewhere in the book.

---

[1] See Field [1972].

[2] This account was taken from Field [1975] especially pp. 382–3.

In 'Reference and Understanding' Putnam argues that the theory of language understanding should be sharply separated from the theory of reference, that is, of the correspondence between words and things and between sentences and states of affairs. For he denies that a theory of language understanding can be construed as knowing the truth conditions of the sentences of a language, it being 'unintelligible what *that* knowledge *in turn* consists in' (p. 129). Our understanding of a language consists in our ability to use it. As an account of use Putnam suggests a probability model for assent to, and dissent from, sentences based on the work of Reichenbach and Carnap. But he is not wedded to such an account; perhaps a Wittgensteinian one would do just as well. Our use of language also enters into our general behaviour and contributes to the success of this (non-linguistic) behaviour. For example, we may say certain things to one another, conduct certain reasonings and manipulate materials in order to successfully build, say, a bridge. But a theory of language use says little of the reference of words to things or of the correspondence of sentences to states of affairs; nor can it explain the contribution that language makes to the success of our general behaviour. Yet there do exist such correspondence relations which are not part of the theory of use. Putnam suggests an analogy. Instructions for the correct use of a tool or instrument (*e.g.* the flipping of a switch to illuminate a room) do not contain an explanation of why following the instructions leads to success. Yet we can all follow the instructions successfully. An explanation of this success would appeal to theories of electricity, circuitry and mechanics which are not part of the instructions. Similarly, the explanation of the success of our general behaviour due to the contribution made by our use of language utilises the correspondences that there are between words and things and between sentences and states of affairs. Thus truth and reference, while not playing as prominent a role in the theory of understanding as those who equate understanding a sentence with knowledge of its truth-conditions would have it, do play a central role in explaining the contribution our use of language makes to the success of our general behaviour. The burden now falls upon showing how this success presupposes truth and reference.

Putnam offers a 'transcendental argument' to show how this success is made possible by the correspondence relations of truth and reference. No complete argument is offered, but there is a programmatic sketch of how, in the simple case of visual perception, we can reliably produce true sentences (*e.g.* 'the rug is green') when certain states of affairs obtain (*e.g.* the rug being green). The argument sketch employs crucially the causal theory of perception and Tarski's familiar truth definition for sentences. Putnam's approach is as programmatic as Field's. But it does suggest a useful alternative account of the role of our notion of reference (and of truth) that is different from Field's physicalist reductive account; it also avoids Field's criticism of Tarski's theory that a theory of reference should not be given a trivial definition in the form of a list. Further, it shows how we can keep both (*a*) Wittgenstein's earlier 'picture' theory of meaning, or a truth-conditional account of meaning, and (*b*) Wittgenstein's later 'use' theory of meaning, or some other 'use' theory. The former will explain the sucess of the latter theory while not being constitutive of that theory.

---

[1] Bracketed page references are to the book under review.

In the John Locke lectures Putnam also charges that Field's approach to a theory of reference is a species of misplaced 'scientific utopianism'. The grounds for this claim arise out of Putnam's views on the relativity of explanation and his comments on the theory of translation. Not every deduction of an explanandum is an explanation; moreover, either explanation has a pragmatic aspect, or it is relative to an 'explanation-space', or it is relative to the interests that there are in putting forward the explanation. (These are not all the same though Putnam sometimes treats them as if they were.) Thus an explanation by Willy Sutton, the bank robber, as to why he robs banks, *viz.* 'because that is where the money is' (and not, say in a shop), is an explanation relative to the interests of Sutton and not, say, a priest or a social reformer who wants to know why Sutton leads a life of crime rather than be law-abiding. Putnam's constraints on the purely deductive model of explanation are important. How do they bear on Quine's theory of radical translation?

Translation of a native's language involves finding a rational explanation of the behaviour, both linguistic and non-linguistic, of the native. To the extent that we can impute to the native our own interests and our own 'explanation-space' then there is no Quinean indeterminacy of translation (and reference). But where two translators from different cultures with different interests and 'explanation-spaces' attempt to translate the same native language there will be indeterminacy in translation. Thus the interest-relativity of explanation and the indeterminacy of translation, Putnam argues, go together. In our everyday practice the translation of a foreign language takes place fairly smoothly despite the difficulties that there seem to be in principle according to the Quinean theory of translation. This practical success seems to have only one explanation: the same interests, thoughts and goals seem to be common throughout humanity. If this is the case, why not build into the theory of translation a theory about this common human nature thereby removing the implicit relativity of translation to our interests, thoughts and goals?

Putnam argues that if we were to render translation scientifically precise in this way then we would have produced an entirely unusable theory of translation to replace the human activity of translation that works very smoothly in practice in the absence of such scientific precision. The goal of producing a scientifically precise translation theory is a species of scientific utopianism. Similarly, argues Putnam, we can criticise Field's reductive physicalist theory of reference as a species of scientific utopianism. It is not that we can show that the Field programme for reference is unattainable. Rather it is not needed because we can provide an alternative account of the role reference (and truth) play that is entirely adequate to our purposes. With these points established Putnam then argues that Mill's claim that 'the backward state of the Moral Sciences can only be remedied by applying to them the methods of Physical Science, duly extended and generalized' (quoted on p. 66) is also a species of scientific utopianism based on an illegitimate extension of a conception of science founded upon physics.

The arguments against scientific utopianism in these various fields turn on the claim that we cannot have a detailed explanatory account of a human being in the way we can have, say, a detailed explanatory account of a hydrogen atom in quantum theory. The reason is not that, like explaining the weather, we have all the laws available (or could in principle have them) but lack all the requisite parameters and find the computations too complex to make. The problem is rather one of 'complexity of structure' (p. 62) in the case of human beings. But what is this? The explanatory models of physics can be applied only if we add the premise 'a human being is a physical system consisting of elementary particles'. But this is to provide an explanation along physicalist lines of the genus physical system and not an explanation of one of the species of this genus whether it be human being, cat, whale, or tree. Putnam rules out as trivial any explanation of the wider genus, physical system, that also counts as an explanation of every species within the genus. Thus, he concludes, physics cannot provide an explanatory model of the natural kind human being. But there is no further argument as to why the triviality requirement is imposed. Any tenacious physicalist would presumably argue that this is part of the very dispute between himself and those who claim that the explanatory models of physics are an aspect of scientific utopianism.

The explanatory models of physics are just part of the conception of scientific method that we have inherited from the dominant science physics. Putnam thinks that inductive logic (p. 70) and Bayesianism (pp. 74–75) are also part of that conception of method. But he nowhere discusses the claims of people such as Popper, Lakatos, Toulmin and Feyerabend who, albeit on very different grounds, argue that we have had a mistaken conception of what are the methods of even physics. Putnam seems to fall into this camp with his suggestion (pp. 72, 73) that unformalised practical knowledge (which he compares to Polanyian 'implicit knowledge') is, along with inductive logic, part of the scientific method of even physics. But his one example of this, a decision as to how well a real open system approximates to an ideal closed system of a theory, is not something which is inconsistent with, say, Popperian methodology. In addition, it is part of Popper's argument in the introduction to *The Poverty of Historicism* that social systems are never like the closed systems of physics and hence historicist methods cannot be applied in the moral sciences. And here historical methods are just those explanatory models of physics extended to social systems and human beings that Putnam has also been critical of. However, Putnam has usefully reminded us of the role practical knowledge could have in the moral sciences and that scientific knowledge is not to be identified with purely theoretical knowledge; but these themes do not receive extensive treatment in the final Lecture where they are introduced.

Putnam has always been a realist, but the kind of realist he is undergoes some transformation in the course of the book. Realism in science is understood as an over-arching empirical hypothesis which involves (at least) the two claims (adopted from an unpublished paper by Richard Boyd) that '(1) Terms in a mature science typically *refer*; (2) The laws of a theory belonging to a mature science are typically approximately *true*' (p. 20). In fact that terms refer and laws are approximately true is part of the explanation of the success of science. Imagine that we interpret a theory instrumentalistically (a species of non-realism) in which the theoretical terms lack reference; suppose also that the theory gives good predictions (within some range). Then the success of an instrumentalistically construed theory lacks an explanation and seems to be miraculous. The atomic theory of matter, for example, may give good predictions when interpreted instrumentalistically but since there are no atoms, protons, electrons *etc.*, which the theory could be about then we lack an explanation of the success of the atomic theory of matter. Here truth and reference explain the success of science in much the same way, noted earlier, that truth and reference explain the success of our general behaviour due to the contribution made to it by our linguistic behaviour. Thus the success of science provides grounds for the realist's general presupposition that the terms of our science do on the whole refer and that its laws are at least approximately true, or true in some limit.

However, it has been argued, especially by Kuhn and Feyerabend, that the history of science shows that successive paradigms, or sufficiently deep theories, are not ontologically convergent in that they are about very different entities, *i.e.* they are referentially incommensurable. On the basis of this history Putnam says that we can make the following meta-induction: 'just as no term used in the science of more than fifty (or whatever) years ago referred, so it will turn out that no such term used now (except maybe observation terms, if there are such) refers' (p. 25). Thus there is a conflict over the convergence of successive theories between the non-realist view of the growth of science of Kuhn and Feyerabend and the realist view of Putnam. To counteract this Putnam supplements his realism with a 'principle of the benefit of doubt' which bids us to assign, from the stance of our current science, the same referent to a term of an earlier theory as possessed by a term of our current theory. Thus Mendel's term 'factor' (or, more strictly, the German term) is given the same referent as our term 'gene', J. J. Thomson's term 'electron' is assigned the same referent as the term 'electron' as used by a person versed in present-day quantum theory, and so on. The detailed workings of this principle are not discussed by Putnam. Presumably no reference is to be assigned to the earlier terms 'celestial sphere' and 'phlogiston'; and there remains a number of highly debatable cases such as the pairs 'impetus' and 'inertia', 'Newtonian mass' and 'relativistic mass' *etc.* In addition it is hard to see how one can preserve from the stance of Keplerian or Galilean astronomy any of the theoretical terms of pre-Copernican astronomy from Eudoxus up to and including even Copernicus himself. However the principle is to be applied in these cases, its general failure would undermine the realist hypothesis while its general success would support it.

But the principle of the benefit of doubt is not an argument against the non-realist view of science. Rather it is a regulative principle which bids us to adopt a convergent view of science when we come to assign referents to the terms of earlier theories from the stance of our own theories. However, Putnam offers an interesting interpretation of the non-realist position in order to re-establish realism. Let us suppose that a scientific theory can be expressed in a language L. Then in the metalanguage we can form Tarski T-sentences of the form:

> (T) 'Venus has carbon dioxide in its atmosphere' is true if and only if Venus has carbon dioxide in its atmosphere.

But, Putnam argues, the formal logic of 'true' and 'refers' as captured by Tarskian semantics leaves underdetermined our concepts of truth and reference. Moreover we normally interpret the logical connectives in the Tarskian theory classically, *i.e.*, realistically. But the 'meaning' of these is also underdetermined by the contexts in which they occur; they could be interpreted intuitionistically, *i.e.* non-realistically. If they are interpreted intuitionistically then in (*T*) we can replace 'true' by 'warrantedly assertible' and add to the right-hand side something like 'provable in *B*' where *B* is our current theory of the planets and chemistry. That is, we can reconstrue (*T*) as:

> (*T.I.*) 'Venus has carbon dioxide in its atmosphere' is warrantedly assertible if and only if 'Venus has carbon dioxide in its atmosphere' is provable in *B*.

And all of this can hold even if there is no Venus and no carbon dioxide because existence itself becomes an intra-theoretic notion. In addition truth becomes intra-theoretic like proof (or whatever expression such as 'warrantedly assertible' takes the role of truth). This is closely allied to the position Kuhn and Feyerabend do adopt on truth and on existence.

However, Putnam re-establishes the realist position by adopting a strategy for 'true' that is akin to the naturalist fallacy argument for 'good'. If one understands the notion of truth realistically and does not construe it non-realistically as in the case of (*T.I.*), then one would accept true modal statements of the following kind: That Venus has carbon dioxide in its atmosphere might not be true even though it follows from our theory *B* that Venus has carbon dioxide in its atmsophere. This is a statement a non-realist who adopts (*T.I.*) cannot accept. By admitting the modal statement the non-realist would be falling in with the realist's concept of truth; for he would be admitting that what is provable within his current theory, and thus warrantedly assertible, might not, after all, be true. That is, it is always open, as it is in the naturalistic fallacy argument for 'good', to claim that even though statement *S* is warrantedly assertible, *S* might nevertheless not be true. To admit modal statements of the above kind into science is tantamount to claiming that truth cannot be the same as warranted assertibility. Thus the thesis of realism bids us to understand truth in a particular way. In addition the thesis is empirical in that it could turn out to be false (*i.e.* there is no convergence in science). At the same time it is a high level claim about the nature of scientific knowledge and an explanation of the success of that knowledge.

In the final essay 'Realism and Reason' Putnam expounds a modified version of his realism which he now calls 'internal realism' as opposed to metaphysical realism which he argues is incoherent. Internal realism retains realism as an empirical hypothesis in which truth and reference have a role in explaining the success of science and the success of our general behaviour that is due to the contribution made by our use of language. Metaphysical realism goes further in claiming that for every correct theory 'there has to *be* a determinate relation of *reference* between terms in [language] *L* and pieces (or sets of pieces) of THE WORLD' (p. 125). On this view THE WORLD is given independently of any theory we may have of it. Moreover, THE WORLD may be such that no theory we could ever produce (even in the ideal ultimate limit of scientific theory as Peirce and Sellars suppose) is a correct representation of it; as Putnam suggests maybe we are really all 'brains in a vat', but the ideal final theory does not represent THE WORLD as this. Thus for the metaphysical realist THE WORLD may be like the Kantian noumenal world about which we can make no intelligible claim as to how it is. Internal realism is supposed to be free of these claims of metaphysical realism because it does not presuppose THE WORLD which is independent of all description; it can perform its explanatory role within the confines of the world as described in a particular way (*i.e.* something akin to Kant's empirical world).

As interesting as these claims are the arguments for them are not as complete or as perspicacious as could be wished. A general argument is provided (pp. 127–9) based on Dummett's non-realist semantics apparently presented in his unpublished William James Lectures. But the account of Dummett's views is too cryptically presented to show the reader why metaphysical realism is untenable. However, this claim is illustrated for the case of a world which is only a finite straight line. We could give an account of this world in three different ways. (1) There are points and line segments which have points as parts. (2) There are no points but only line segments which may be of an irrational length. The talk of points in (1) now becomes highly derived talk of convergent sets of line segments. (3) There are no points and only line segments of a rational length. Points and irrational line segments are now logical constructions out of rational line segments. We now find that we can no longer speak of this straight line world independently of any of the three accounts given. There is no fact of the matter as to whether there are points or only highly derived talk of line segments. Thus theory (1) of the world attributes entities, points, to the world that theories (2) and (3) do not. There is not even any fact of the matter as to the cardinality of the world; for (1) and (2) there are non-denumerably many objects while for (3) there are only denumerably many objects.

But what of our talk in our world which contains cows when we English speakers use the term 'cow'? Is there no fact of the matter that there are cows out there in the world independent of any of our representations of them? We are familiar enough with Quine's claims about rabbits. But it is not clear how Putnam's ruminations on pages 135–7 make a case for cows akin to that he makes for the world of the single line. We are invited to consider the sentence '"cow" refers to cows' which is trivially true by virtue of our understanding of reference and of the quotation convention. What the last word in this sentence refers to, Putnam alleges, can only be discussed within our theory; to assume that it can be discussed independently of any theory is to assume metaphysical realism, *i.e.* that there is a theory-independent fact of the matter as to what a term in a theory refers to. In addition, since he has separated the theory of the understanding (*i.e.* use) of language from the truth-conditional and reference-conditional account, Putnam claims that the term 'cow' can be used in our language and its use will have already been explained by our theory of understanding. We are then free to use the term 'cow' even in explaining what 'cow' refers to. But to question this is to question just how far we can construct a theory of understanding (*i.e.* use) independently of any contribution that things like cows that are out there in the world might make to such a theory.

ROBERT NOLA
University of Auckland

REFERENCES

FIELD, H. [1972]: 'Tarski's Theory of Truth', *The Journal of Philosophy*, 69.
FIELD, H. [1975]: 'Conventionalism and Instrumentalism in Semantics', *Nous*, 9.

HUTCHISON, T. W. [1977]: *Knowledge and Ignorance in Economics*. Oxford: Basil Blackwell. £7.00. Pp. x + 186.

The purpose of this book is to deflate our expectations of what economics can achieve. T. W. Hutchison argues that economic theory is unable to issue in scientific predictions. Hutchison follows Popper in holding that for a prediction to be scientific it must be deduced from a corroborated universal generalisation plus some independently established initial conditions. The trouble with economic theory, he then argues, is that it lacks any serious such generalisations.

However, Hutchison does not hold that there is no possibility whatsoever of predicting economic events. Even if *scientific* predictions based on universal generalisations are out, there is still room, he argues, for predictions of a somewhat less exalted status: namely, predictions based on the informed extrapolation of trends.

Hutchison sees this as a matter of induction versus deduction. Economists should avoid 'an excessive or exclusivist pro-deductive, and anti-inductive approach'. Economic forecasting should be based on the 'extrapolation of trends by a kind of induction' and so 'induction seems to be much more centrally involved than the hypothetical-deductivist methodology allows (distilled as it has been mainly from the history of the more "advanced" natural sciences)' (p. 23).

There are a number of puzzles about this initial discussion of economic prediction (which takes up most of the first quarter of the book). To start with, it is difficult to know what to make of the contrast between 'inductive' trend-based predictions and 'deductive' law-based ones. Hutchison quotes with approval Popper's observation that '. . . *trends are not laws*. A statement asserting the existence of a trend is existential, not universal' ([1960], p. 115). But he does not face up to the apparent consequence that getting from a trend to a prediction requires at least the implicit postulation of some universal generalisation, and so leaves us in the dark about the real substance of his suggested contrast. It is true that on the laws-and-initial-conditions analysis we arrive at a scientific prediction by *deducing* it from the relevant law and conditions. But that is the easy part of law-based predicting. The hard bit is substantiating the law in the first place. And the philosophical consensus would be that this cannot but involve an inductive inference from the kind of finite pattern which comprises a trend. Of course Popperians deny that the acceptance of laws requires any such inductive support. But this is no help to Hutchison's contrast, for if the Popperian arguments are right they tell against *any* kind of inductive reasoning, including any proposed mode of going straight from trend to predictions. Then there are those who accept inductive reasoning, but argue, like Mill, that this is best conceived of as a matter of going straight from particular judgments to particular judgments, without the aid of any mediating laws. However, there is no indication that Hutchsion is thinking on this level. And in any case this line of thought is scarcely going to give us the desired differentiation between prediction in economics and prediction in 'the more "advanced" natural sciences', for it is equally applicable to both areas.

Still, even if it is not a matter of induction instead of deduction, Hutchison is surely right to want some room for predictions which are less than deductions from exceptionless universal laws but more than mere guess-work. But how are such predictions possible? One idea might be this. Suppose we take some observed trend to be due to some (unknown) law and certain (unknown) standing initial conditions. Will it not then sometimes be reasonable to predict the continuation of the trend, on the supposition that the relevant initial conditions will continue to stand? But this is full of holes. What will justify the underlying assumption that our trend is the manifestation of some law or other, rather than mere happenstance? And what makes the crucial supposition, that the relevant initial conditions will persist, better than a guess, if it is something less than a law-based deduction from certain other features of the situation?

It is surprising that Hutchison ignores the possibility of *probabilistic* prediction almost entirely. On the couple of occasions he does mention it he dismisses it immediately. But it is surely here, if anywhere, that we will find something between unqualified prognosis and pure guesswork. Certainly our attitude to many trends has a probabilistic dimension: instead of taking them to be the fixed outcome of some deterministic law plus constant initial conditions, we see them as manifesting a random variation around some basic pattern. Most frequent is the familiar linear regression model in which one factor (the dependent variable) is taken to be a linear function of certain other factors (the independent variables) plus some random contribution. If we like, the random contribution can be thought of as due to the influence of certain further unknown factors. But now that we are thinking probabilistically we can use statistical techniques to estimate the probability distribution of the random contribution, even if we remain ignorant of exactly what the unknown factors responsible for it are. And then, given such an estimate, together with our assumptions about the underlying function relating our dependent variable to any explicitly considered independent variables, we will be able to predict the probability of the dependent variable falling in any given interval.

Such probabilistic predictions will still depend on universal generalisations. But these will now be statistical, not deterministic, laws. Of course we cannot take statistical laws for granted any more than deterministic ones. Even if we pass over the obscurity of their logical form, and ignore the special problems that arise for the logic of testing statistical hypotheses, we are still left with the general inductive problem: does the pattern we have discerned in our finite experience indeed manifest some universal truth? And this last question will arise both for the postulated underlying function relating the dependent to the explicitly considered independent factors and for the postulated distribution summarising the contribution of the unknown randomly varying influences. Often we will be rather more confident of the former than the latter ('one can extrapolate slopes, but not correlations'). Now, it might be argued that in economics we should never be confident on any such matters—that economic reality is too complex even for statistical generalisations, let alone deterministic ones. But Hutchison does not argue this. This seems especially surprising in view of the huge amount of recent econometric effort put into the elaboration of probabilistic models and the development of statistical techniques for assessing their accuracy. Perhaps this effort is all misplaced. But if so one would like to be shown why—it is certainly not something which follows straight from the admitted implausibility of any serious deterministic laws in economics.

In his third chapter Hutchison considers the methodological writings of Popper, Kuhn, Ravetz and Lakatos. In general he is suspicious of attempts to teach economists lessons learnt from the harder sciences. And in particular he feels that any current enthusiasm for epistemological tolerance should be resisted in economic circles. He quotes Ravetz's view that the practitioners of the more mature sciences have been socialised into automatic conformity to stringent methodological standards. The moral, for Hutchison, is that in economics and other soft subjects, where the institutionalisation of methodological stringency has not yet occurred, there is a need to *enforce* appropriate methodological standards.

In further support of this view Hutchison offers the following quote from Lakatos:

> While Polyanite academic autonomy should be defended for departments of theoretical physics, it must not be tolerated, say, in institutes of computerised social astrology, science planning or social imagistics. ([1971], p. 121).

But this is surely something of a misuse of Lakatos. Lakatos is not, in the article referred to, arguing that immature sciences should be subject to methodological controls from which the hard sciences can somehow be excused. His concern is rather the prior one of finding a basis for deciding between alternative methodologies of science, and his suggestion here is that this can up to a point be done by distilling principles from the actual practice of such exemplary disciplines as theoretical physics. But of course once such principles have been so distilled they will apply to hard and soft sciences alike. There is nothing in what Lakatos says to support Hutchison's contention that in economics we need to guard against the kind of methodological laxity that is all very well for physics.

More generally, Hutchison seems to miss the point of recent arguments for epistemological toleration. As he presents it, such toleration is a luxury that is harmless (indeed irrelevant) in a community already conditioned into strict falsificationism. He quite fails to confront the thesis that some degree of theoretical tenacity in the face of apparent falsification is actually an essential prerequisite of maturity in a science.

Even so, that theoretical tenacity is a necessary condition of scientific maturity does not mean it will be sufficient. Perhaps the complexity of its subject matter means that economics is condemned to immaturity in any case. Hutchison argues along these lines in Chapter 4, '"Crisis" in the Seventies: the Crisis of Abstraction'. Here the taret is the elaboration of abstruse mathematical models which do not even pretend to represent the real economic world. Hutchison considers Joan Robinson's 'optimistic' defence of such theoretical abstraction: she suggests that we can reasonably hope that in time we will be able to modify the assumptions of our 'ideal' models in a realistic direction. His response is to complain that there has been almost no effort on the part of theoretical economists to so modify their models; and he suggests further that if any such attempt were made it would inevitably be frustrated by the shifting complexity of economic reality. Economists would do better to abandon grandiose theoretical aspirations altogether and concentrate instead on informed analyses of particular historical situations.

Hutchison is only too justified in his complaint that few theoretical economists seem to care about the failure of their 'ideal type' models to grip onto reality. But is any attempt to remedy this failure necessarily a waste of effort? Hutchison's pessimism on this point seems to derive in part from his blindness to the possibility of probabilistic theorising. One obvious way of making abstract models more realistic would be to replace deterministic assumptions about what would happen in ideal conditions with probabilistic assumptions about what would happen in various kinds of actual situation. Consider, however, Hutchison's argument against the possibility of modern general equilibrium theory (which incorporates an assumption of perfect knowledge) ever becoming empirically applicable: 'It is bound to be more or less arbitrary, . . ., to assume any one particular form of incorrect expectations or ignorance, as a general case' (p. 80). Indeed. But it does not at all follow that we cannot reasonably make probabilistic assumptions of some kind about the range of expectations that will be found in given economic conditions.

A lengthy appendix to the main text reports on the debate on devaluation amongst economists in Britain between 1964 and 1969. Hutchison shows clearly that, contrary to the later pronouncements of some economists, disagreement on the merits of devaluation was by no means confined to politicians alone, but was also widely present amongst professional economists. As an illustration of the impossibility of economic prediction, however, this devaluation example is less convincing. There is little evidence, in the many quotes Hutchison produces, of any divergence on the basic principles of exchange rate theory. The difficulty with devaluation is rather that its eventual effects on the balance of payments and other indices of economic well-being are extremely sensitive to the precise values of various economic parameters, including parameters over which the government can exercise some degree of control. Apart from anything else, uncertainty about what back-up policies, if any, will be instituted make it hazardous to predict even the direction of the relevant effects of a given devaluation. But it scarecely follows that the deliverances of economic prognosticators will in all cases be equally equivocal. One might as well argue against the overall predictive utility of physics on the grounds that physicists disagree on whether or not the building of nuclear reactors will lead to large-scale radiation poisoning.

A number of underlying themes run through the book. There are various pertinent observations about recent changes in the institutional roles played by economists, both inside and outside academic institutions (though, surprisingly, there is no discussion of the significance here of post-war advances in computer technology). On a more 'internal' level, Hutchison in a number of places criticises economists' (especially British economists') lack of historical awareness. He suggests that 'a more empirical, historical and institutional emphasis' (p. 96) would do much to counteract the dangers of theoretical abstractionism. If Hutchison is recommending here that economic theorists should consider 'social' variables along with strictly 'economic' ones, then he is certainly on strong ground. The idea that a set of economic variables alone can ever form an explanatorily complete set, even on a probabilistic level, has little prior plausibility. (That this idea seems to be an implicit presupposition of much economic theorising testifies to nothing so much as the power of demarcation disputes between different academic disciplines to distort intellectual developments.) However, it is not always clear that what Hutchison is recommending is simply that economists should include social variables in their generalising. As often what he seems to have in mind is rather that they should give up generalising altogether and concentrate instead on the spotting of trends in particular historical episodes. We have already seen that *this* recommendation, if taken seriously, would in the end turn economic prediction into nothing more than random guesswork.

Reflected in the title of the book is what Hutchison takes to be his central message—that it is crucially important for us to recognise the weaknesses and limitations of the prognoses on which our economic policy decisions are based. But even here it is often unclear exactly what Hutchison is driving at. Sometimes his point seems merely to be that since economic prognoses are fallible we should be ready for the possibility that our policies will fail to produce the intended results. One can scarcely argue with this. And it is indeed worth reminding ourselves that the ineffectiveness of some economic policy does not necessarily impugn the decision-making competence of those who instituted it. But sometimes Hutchison seems to have a rather stronger moral in mind: namely that an awareness of our prognostic limitations should actually affect our choices of economic policies, and in particular that it should dissuade governments from active intervention in the economic system. Thus, for instance, in the appendix on the devaluation debate Hutchison responds to a contribution which disclaims any 'precise knowledge of the quantitative effects of a given devaluation on the U.K. balance of payments' as follows:

> But if an academic, or anyone else, can only discuss the issue in non-quantitative terms of 'principle', with no precise quantitative estimates, *it might seem, perhaps, that he could go little further in terms of support than to throw out the question, when balance of payments difficulties arise, of 'What about devaluation?'* (p. 109, his italics).

There are indications in various places that Hutchison is in sympathy with the recent 'anti-Keynsian' opposition to government economic intervention. (Thus, although he snipes at Joan Robinson for her 'optimistic' defence of theoretical abstraction, he does not mention Milton Friedman's far better known, and much less coherent, advocacy of 'unrealistic' economic models.) Now, perhaps there are good arguments for monetarist anti-interventionism. (Whether the suggested policies can seriously be considered 'non-intervention', rather than intervention of a quite specific kind, is another matter.) But the imperfection of our economic knowledge in itself provides no reason whatsoever for anti-interventionism. Even if our understanding of economic matters is less than perfect, it's all we've got. What alternative is there, when we are faced with some economic decision, but to judge as best we can the likely results of the alternative policies available, and to choose accordingly? That some policy might be describable as distinctly 'non-interventionist' does not entitle it to any special treatment in this context—its expected consequences will still need to be estimated in just the same way as those of the other alternatives.

Here again it seems that Hutchison might have been led astray by his inattention to the possibility of probabilistic prediction. For the sake of argument, suppose that we only took sure things into account in our decision-making, and that we simply ignored as irrelevant any possible consequences of our actions whose occurrence or non-occurrence we were less than certain of. Now one thing we are in general sure of is the direct cost in terms of time and resources of administrating some policy and getting the relevant people to conform to it. By contrast the intended benefits of any economic policy will amost invariably be indirect and to some degree uncertain. So under the suggested decision procedure the intended advantages of the various alternative policies would count for nothing, and we would indeed be directed to that choice which minimised the direct administrative costs of intervention. But clearly this would be a quite absurd decision procedure. Not only would it prevent us ever taking any active risks in the hope of improving our economic position, but it would also force us always to ignore any bad effects that were likely, but short of certain, to result from non-intervention. Of course it is true that the more sure we are of a given result, the more it should weigh for or against the relevant policy in our decision-making. And this does mean that in general the undesirability of administrative costs should receive proportionately greater weight than the attractiveness of intended results. But that certainties weigh more than probabilities by no means precludes the possibility that on occasion interventionist policies might still be the best bet. To suggest otherwise is plausible only if we make the mistake of discounting probabilities altogether.

There is one further sense in wich the imperfection of our economic understanding might be thought to provide a reason for economic inaction. Quite apart from issues of probabilities versus certainties, there is the question of whether our economic views, of whichever kind, are adequately supported by the available evidence. If we were to decide that certain of our economic beliefs were inadequate in this specifically epistemological sense, then the appropriate response might well be to postpone any further action until more evidence has been gathered. But it is unlikely that this is what Hutchison has in mind. For he does not think that the deficiencies in our economic understanding are something that can be remedied by more thorough research and investigation—on the contrary, his view is that the imperfections of economic thinking follow from the nature of the subject, and that we should simply come to terms with this. But incurable epistemological doubts have no practical consequences. Even if we do recognise that the theories on which we anticipate the economic future are less fully evidenced than we might like, our inability to get better theories means that we will have no alternative but to settle for the guide we have, and so we will end up making just the same economic decision as we would if we had perfect evidence for those theories. It is true that the more we doubt the evidential warrant for our theories the more we will expect our decisions to turn out *post hoc* to be unsuccessful. But that is a different point—to have a prior doubt about a decision is not yet to have a reason for making a different one.

DAVID PAPINEAU
Macquarie University

REFERENCES

Popper, K. R. [1960]: *The Poverty of Historicism*, 2nd ed. London: Routledge and Kegan Paul.

Lakatos, I. [1971]: 'The History of Science and its Rational Reconstructions' in R. S. Cohen and R. C. Buck (*eds.*): *Boston Studies in the Philosophy of Science*, **8**.

Williams, B. [1978]: *Descartes: the Project of Pure Enquiry*. Sussex: The Harvester Press. £10.50. Pp. 320.

Curley, E. [1978]: *Descartes Against the Skeptics*. Oxford: Basil Blackwell. £9.50. Pp. xviii + 242.

Each of these books contains one chapter on each of the familiar fragments of Descartes's reasoning: one chapter on the *cogito*, one on Descartes's arguments for God's existence, one on his theory of body, one on his theory of mind. Each book also reports an attempt to describe the project into which these fragments fit, to explain why Descartes believed that the project would lead him to perfect certainty, and to say why Descartes valued certainty so highly that he untertook the project despite its difficulty.

The question on which Williams focuses is 'Why did Descartes search for certainty rather than just for truth?'. To answer this question, Williams describes a project which he calls 'the project of pure enquiry'. He then contends both that the project was Descartes's and that, in the context of the project, the search for truth necessarily becomes the search for certainty. Since Williams describes the project of pure enquiry in twentieth century language, and since he argues that the project is one which a twentieth century person might reasonably undertake, he believes that he has performed 'a rational reconstruction of Descartes's thought, where the rationality of the construction is essentially and undisguisedly conceived in a contemporary style' (p. 10).

As Williams defines 'pure enquiry', to engage in the project of pure enquiry is to make an uncompromising attempt to raise one's own 'truth ratio' to unity—one's truth ratio being the number of his true beliefs divided by the total number of his beliefs, true or false.According to Williams, a person will succeed in this project if and only if he adopts a method for acquiring beliefs which satisfies two conditions: First, the method must be 'error proof' (p. 48); that is (in some sense of 'possible') it must not be possible to acquire a false belief as a result of a proper application of the method. Second, the method must be 'epistemically effective' (pp. 48–49); that is (in some sense of 'possible') it must be possible for the enquirer to tell, as he proceeds, whether or not he is applying the method properly. Since Williams believes that a person who uses such a method will *ipso facto* become certain of the truth of each of the beliefs which he acquires, he concludes that the pure enquirer's search for truth necessarily becomes the search for certainty: 'What is needed [by the pure enquirer] is an error proof method which is epistemically effective. *This comes to* the requirement that the beliefs which the method generates be *certain*' (p. 48; emphasis mine).

To complete his explanation of Descartes's having searched for certainty, Williams argues that Descartes was a pure enquirer. Williams' strategy here is to point to facts about the *Meditations* which are puzzling initially but which cease to be puzzling when we suppose that Descartes was a pure enquirer. Williams points, for example, to Descartes's attempt in *Meditation I* to rid himself of all his previous beliefs; according to Williams, Descartes's having been a pure enquirer would explain his having made this attempt, since the pure enquirer's only hope for success lies in his controlling his method for *acquiring* beliefs (pp. 39, 45, 49). Williams also points to Descartes's attempt to isolate himself from all practical concerns; according to Williams, Descartes's having been a pure enquirer would also explain his having made this attempt, since the pure enquirer's decision *uncompromisingly* to raise his truth ratio to unity threatens so to impoverish his system of beliefs that the system no longer provides guides for practical action (pp. 34, 46–47).

Is Williams's analysis here helpful? Williams does not offer precise or explicit definitions of such crucial terms as 'error proof' and 'epistemically effective', and, since he offers no reason for thinking that a person's set of beliefs is countable and finite, his introduction of the term 'truth ratio' is not fully justified. In consequence, as Williams defines the term 'pure enquiry', to say that a person is a pure enquirer is just to say that the person is making an uncompromising attempt to rid his system of beliefs of error and falsity. Since it is obvious to every intelligent reader of the *Meditations* that Descartes made such an attempt, the identification of Descartes as a pure enquirer does not seem to me to advance our understanding of Descartes's project.

Williams's reply might be that, if we talk about the pure enquirer rather than about the historical Descartes, we can isolate the philosophically interesting aspect of Descartes's project from the historical irrelevancies, thereby achieving clarity. But clarity is not forthcoming. Williams's explanation of the pure enquirer's desire for certainty rests on the contention that a person who adopts an 'error proof' and 'epistemically effective' method will *ipso facto* become certain of the truth of the beliefs which he acquires, but Descartes himself would have rejected this contention. On Descartes's view, a person who has even the slightest suspicion that there may not be a veracious God ought to have the suspicion that his mind is poorly designed (*Meditation I*; HR I, pp. 147–8; AT VII, pp. 21–22),[1] and a person who has even the slightest suspicion that his mind is poorly designed ought not to be perfectly certain of the truth of the beliefs which he acquires. On these gounds, Descartes explicitly contended that a person cannot be perfectly certain of anything unless he is certain that there is a veracious God (*Meditation III*; HR I, pp. 158–9; AT VII, p. 36) and, in particular, that an atheist mathematician cannot be perfectly certain of the truth of the theorems which he proves (*Replies II*; HR II, p. 39; AT VII, p. 141). In other words, Descartes contended that, for a person to be certain of the truth of a belief, it is *not* sufficient that the person have acquired the belief by the proper application of an 'error proof' and 'epistemically effective' method; for a person to be perfectly certain of the truth of a belief, it is also necessary that he be certain that his mind is a reliable instrument for discerning truth.

Of course, there is a problem here: The view that a person cannot be certain of anything unless he is certain that his mind is a reliable instrument for discerning truth seems to entail that a person who is not certain that mind is a reliable instrument for discerning truth cannot *become* certain of anything. So it seemed, for example, to Sextus Empiricus (in *Against the Logicians*) and to Montaigne (in *Apology for Ramond Sebond*). From the premise that a person cannot be certain of anything unless he is certain that his mind is a reliable instrument for discerning truth, such sceptics infer that the search for certainty must either go around in circles or regress *in infinitum*. Their reasoning is that, if we once are *un*certain that our minds are reliable instruments, we must become certain that they are reliable instruments before we can become certain that they are reliable instruments.

---

[1] References beginning with 'HR' are to Haldane, E. S. and Ross, G. R. T. (trans.) *The Philosophical Works of Descartes* (1970 edition); those beginning with 'AT' are to Adam, C. and Tannery, P. (*eds.*) *Oeuvres de Descartes* (1904 edition).

As Curley shows in the first chapter of *Descartes Against the Skeptics*, Descartes was aware of this sceptical argument. Yet, in the *Meditations*, Descartes accepts the premise of the argument without accepting its conclusion: Although he admits (in *Meditation I*) that he is not certain that his mind is a reliable instrument, and although he admits (in *Meditation III*) that he cannot be certain of anything until he is certain that his mind is a reliable instrument, Descartes believes that he *can* attain perfect certainty. That is, Descartes's attempts to attain a state which the *sceptics* would call a state of certainty. In contrast, since the only certainty that the pure enquirer wants is the certainty that comes from adopting an 'error proof' and 'epistemically effective' method, and since a person can adopt such a method without being certain that his mind is a reliable instrument for discerning truth, the pure enquirer does not attempt to attain a state which the *sceptics* would call a state of certainty. And this shows, I suggest, that the project of pure enquiry is not a complete reconstruction of Descartes's project: Unlike the pure enquirer, Descartes wants to refute scepticism.

Descartes's relationship to the sceptics is, of course, the focus of Curley's book. Curley's view, as he states it, is that Descartes replaced the sceptics concept of proof with a new concept, thereby avoiding sceptical problems (p. 118).

At the heart of Curley's discussion of these issues is an analysis of the phrase 'valid grounds for doubting' which can be summarised as follows: a person X has valid grounds for doubting a proposition p if and only if there is an hypothesis q such that (i) either q is incompatible with p or q is incompatible with some other proposition, r, which provides X with his 'basis for affirming' p, (ii) X can have some suspicion, however slight, that q may be true, and (iii) if q were true, it would explain how X acquired the false belief that p is true (p. 122). Curley does not explicitly say what he means by the phrase 'basis for affirming'. Nor does he explicitly say how an hypothesis can explain our acquiring a *false* belief when it is compatable with the truth of the proposition that is believed. But Curley's discussion does suggest an example: the proposition 'All that I am compelled to affirm is true' is one's basis for affirming the simple and evident proposition '$2+2 = 4$' (which, on Descartes's theory, is a proposition which one is compelled to affirm), and the proposition 'My mind is so poorly designed that the propositions which I am compelled to affirm are false' therefore satisfies conditions (i) and (iii) without itself being incompatible with '$2+2 = 4$'.

Examples of this sort lead Curley to a plausible view concerning the relationship of those passages in which Descartes says that he *cannot* doubt simple and evident propositions (*Replies II*: HR II, p. 42; AT VII, pp. 145–6) to those passages in which he says that he *has* valid grounds for doubting them

(*Meditation III*; HR I, pp. 158–9; AT VII, p. 36).[1] On Curley's theory, there is an hypothesis *q*—namely, 'My mind is so poorly designed that some of the propositions which I am compelled to affirm are false'—which is such that (i) *q* is compatible with Descartes's 'basis for affirming' each simple and evident proposition, (ii) Descartes has some slight suspicion (before *Meditation III*) that *q* may be true, and (iii) if *q* were true, it would explain Descartes's having misjudged the simple and evident propositions. Hence, Curley's analysis of 'valid grounds for doubting' entails that (before *Meditation III*) Descartes has valid grounds for doubting the simple and evident propositions which he is compelled to affirm and which he therefore cannot doubt. And, on Curley's view, Descartes has these grounds even though—or, rather, precisely because—he cannot doubt that these propositions are true.[2]

Curley's primary use of his analysis of 'valid grounds for doubting' comes, however, in the course of his attempt to describe the relationship of Descartes's philosophy to traditional scepticism. According to Curley, Descartes wanted to refute the sceptical view that, since we can never *prove* our initial axioms, our initial choice of axioms is 'arbitrary' (p. 84). As Curley sees it, Descartes attempts to refute this view by 'implicitly redefining the concept of proof' (p. 118) so as to allow that one can prove even his initial axioms. When explaining Descartes's new concept of proof, Curley attributes to Descartes the view that a proposition *p* is proved for a person *X* if and only if two conditions are satisfied: The first is that *X* is compelled to affirm *p* (either because *p* is simple and evident or because *X* has in mind an assent compelling argument for *p*), and the second is that *X* has no valid grounds for doubting *p*. In this sense of 'proved', says Curley, Descartes has proved the simple and evident propositions which are the initial axioms of his system, and he therefore has avoided the sceptics' complaint that his initial axioms are 'arbitrary'.

Curley does not provide us with a description of the sceptical concept of proof which he supposes Descartes to be 'redefining'; he does not say what it is to 'redefine' a concept; he does not provide us with a detailed or scholarly account of the sceptical argument that he takes Descartes to be refuting; and he does not provide textual support for his contention that Descartes was interested in the concept of proof. In fact, by saying that Descartes *implicitly* redefines the concept of proof, Curley implicitly admits that he is not fully justified in attributing to Descartes a concept of proof according to which a person can prove a proposition *p* without even constructing an argument for *p*'s truth. Curley therefore offers his readers an alternative: "If you find this use of 'proved' too strange", he writes, "you may treat this as an account of the normative sense of indubitability" (p. 123). But 'indubitability' is not quite the right term either: Curley wants a term which applies to a simple and evident proposition, not just because we have no reason to doubt its truth, but also because we have a strong conviction that it is true. I therefore suggest that the term for which Curley searches is 'certain' and that the definition which Curley unsuccessfully attempts to articulate is that a person is *certain* of the truth of a proposition *p* if and only if he is compelled to affirm *p* and he has no valid grounds for doubting *p*.

---

[1] Williams attempts to resolve this tension by saying that, although Descartes cannot doubt one of these propositions while he has it in mind, he can doubt it when he does not have it in mind (p. 186). But there are two problems here: first, when Descartes says that he has grounds for doubting simple and evident propositions, he *does* call several of these propositions to mind (*Meditation III*; HR I, pp. 158–9; AT VII, p. 36). Second, Descartes explicitly contends that a person cannot doubt a proposition unless he has it in mind (*Replies II*; HR II, p. 42; AT VII, pp. 145–6).

[2] Compare my [1977], pp. 201–2.

If Curley were to accept this suggestion, however, he would need fundamentally to revise his views concerning the relationship of Descartes's philosophy to traditional scepticism. From the view that a person is certain of the truth of a proposition only if he has no valid grounds for doubting it, it seems to follow (on Curley's analysis of 'valid grounds for doubting') that a person can become certain of the truth of a proposition only if he is first certain of the reliability of his methods for acquiring beliefs or, more generally, that a person can become certain of the truth of proposition only if he is first certain of the truth of another proposition. But the view that a person cannot become certain of something unless he is first certain of something else is the premise of Sextus's and Montaigne's argument to the conclusion that certainty is unattainable. If Curley were to admit that he uses the terms 'proved' and 'indubitable' as surrogates for the Cartesian term 'certain', he therefore would need to abandon the view that Descartes avoids sceptical problems by 'redefining' the sceptics concepts: Even if Descartes's concept of certainty does differ from the sceptics', his concept of certainty seems to generate, rather than to avoid, the sceptics' infinite regress.

On the other hand, there may be benefits which follow from viewing Curley's terms 'proved' and 'indubitable' as surrogates for the Cartesian term 'certain'. When we view the terms in this way, Curley's discussion of Descartes's project suggest an answer to Williams's question: 'Why did Descartes search for certainty rather than just for truth?'. On the definition of 'certain' which I am suggesting, a person *X* is certain of the truth of a proposition *p* only if *X* has no valid grounds for doubting *p*. But, on Curley's analysis of 'valid grounds for doubting', *X* has no valid grounds for doubting *p* only if, for each hypothesis *q* which is incompatable with *p*, *X* cannot have even the slightest suspicion that *q* is true. Therefore, on this definition of 'certain', if *X* is certain of the truth of *p*, *X* cannot be convinced by argument or evidence that *p* is false. It follows that, if there were a set of axioms of whose truth everyone in the scientific community were certain, these axioms would provide a 'firm and secure foundation for the sciences'—a foundation which would be unlike the foundation of Aristotlean science in that argument and evidence would never force us to abandon it.

RONALD RUBIN
Pitzer College

REFERENCE

Rubin, R. [1977]: 'Descartes's Validation of Clear and Distinct Apprehension', *Philosophical Review*, **86**, pp. 197–208.

Brit. J. Phil. Sci. 31 (1980), 109 *Printed in Great Britain*

Notes to Contributors

Articles must conform to the following conventions:

1. Three copies of all submitted articles should be sent typed on one side only with wide margins and double-spaced. They should be accompanied by abstracts of about 100 words.
2. Type of print (capitals, italic, bold, roman *etc.*) should be indicated in the margin.
3. Any section headings should be in small capitals with section numbers in bold face print. These headings should be listed (in italics) in a contents list at the beginning of the article.
4. Long quotations should be indicated in the margin and are printed in small print without quotation marks. Short passages quoted within the text should be enclosed by single quotation marks; double quotation marks should be used *only* for quotations within quotations and for "scare quotes".
5. English letters used as mathematical or logical symbols should be in italics (vectors in bold). Foreign letters and any special symbols should be indicated clearly in the margin. All formulae should be in such a form as to minimise the number of printed lines (so, *e.g.* dy/dx is to be preferred to $\frac{dy}{dx}$ and $(1-v^2/c^2)^{-\frac{1}{2}}$ to

   $$\frac{1}{1-\frac{v^2}{c^2})}.$$

   For further advice authors are referred to Chaundy *et al.* [1964]: *The Printing of Mathematics*, especially pp. 100–1.
6. Tables and illustrations must be professionally drawn on separate sheets with their required place in the text clearly indicated.
7. A list of 'References' in alphabetical order should be placed on a separate sheet. *Layout and punctuation must follow present* Journal *practice, e.g.*

   RUSSELL, B. A. W. [1903]: *Principles of Mathematics*. London: George Allen & Unwin.

   ZAHAR, E. G. [1973]: 'Why Did Einstein's Programme Supersede Lorentz's (I)?' *The British Journal for the Philosophy of Science*, **24**, pp. 95–123.

   Simple citations should then be made, *not* in footnotes, but in the text by naming the book using the author's name and the book's date of publication. (*E.g.* 'See Russell [1903]'. or 'As Russell points out in his [1903] . . .').
8. Footnotes, which should be as few as possible, should be typed separately, double-spaced at the end of the text. They should be referred to by consecutive numbering.

Offprints of articles may be ordered when corrected galley proofs are returned. Fifty are supplied free; the price of extra copies is given on the order form.

PACIFIC
PHILOSOPHICAL
Quarterly
(formerly: *The Personalist*)
Vol. 61, Nos. 1 & 2 January–April 1980

Behaviorism, Language and Meaning	Wilfrid Sellars
Sellars on Behaviorism, Language and Meaning	W. V. Quine
The Logic of Believing	Roderick M. Chisholm
Probabilistic Causality	Wesley C. Salmon
Causal Laws and Effective Strategies	Nancy Cartwright
Truth as a Mode of Evaluation	Brian Ellis
Parapsychology: Science or Pseudo-Science?	Antony Flew
Utilitarianism and Generalized Benevolence	J. J. C. Smart
Legal Moralism and Freefloating Evils	Joel Feinberg
Categories and Spinoza's Attributes	Virgil C. Aldrich
Merleau-Ponty's "Ontology of the Visible": Some Exegetical and Critical Comments	Frederick A. Olafson

Editors: John Hospers, Brian Loar, Bas C. van Fraassen
Managing Editor: Raymond R. Neal
Double Issue: $7.50
Annual Subscription: Individuals, $12 / Institutions, $16 / Students, $9

SCHOOL OF PHILOSOPHY
University of Southern California
Los Angeles, CA 90007 U.S.A.

NOTRE DAME JOURNAL OF FORMAL LOGIC

V. Frederick Rickey, *Editor-in-Chief*

Michele Daugherty, *Business Manager*

The Journal publishes articles in all areas of philosophical and mathematical logic, as well as historical and expository articles in these fields. Works accessible and of interest to both philosophers and mathematicians are especially welcome.

Articles for publication and all editorial correspondence should be sent to the *Editor, P.O. Box 5, Notre Dame, Indiana 46556, U.S.A.*

Published quarterly by the University of Notre Dame in January, April, July, and October of each year. Effective 1 July 1978 and beginning with Volume 20 (1979), subscriptions are $27.50 per year for institutions and $20.00 for individuals. Back volumes are $20.00 each for Volumes 1-10, and $35.00 for Volumes 11-19. Orders for subscriptions and back volumes (accompanied by payment), and changes of address should be sent to the *Business Manager, P.O. Box 5, Notre Dame, Indiana 46556, U.S.A.*

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

DECEMBER 1979, VOLUME XL No. 2

CONTENTS

I. ARTICLES
Referring, by A. P. MARTINICH
Rational Motivation, by HILLIARD ARONOVITCH
Husserl and Heidegger, by H. PIETERSMA
Heidegger and the Modes of World-Disclosure, by SANDRA LEE BARTKY
Criteria of Creativity, by CARL R. HAUSMAN
Causal Efficacy and Causal Explanation, by DOROTHY WALSH
On the Origin of Space Perception, by ALFRED POLITZ
Why is there Something Rather than Nothing? by D. GOLDSTICK
Wittgenstein and Metaphor, by JERRY H. GILL

II. DISCUSSIONS

III. REVIEWS

IV. NOTES AND NEWS, RECENT PUBLICATIONS

Subscription prices $10.00 per year for individuals; $12.00 per year for institutions. Separate issues, $2.50 and $3.00, respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed, stamped envelope, or equivalent postage.
A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo
320 Baldy Hall, Buffalo, New York 14260

MIND

Vol LXXXIX *No 353* *January 1980*

A quarterly review of philosophy
edited by D.W. Hamlyn with the assistance of S.D. Guttenplan

Articles:
More Thought on Thought and Talk: J. BISHOP
Tractatus: Pluralism or Monism?: M. GLOUBERMAN
Epistemic Probabilities: F. W. DAUER
Moral Symptoms: N. RICHARDS
The Expression of Emotion in Music: S. DAVIES

Discussions
Digging Deeper into Determinism: D. LOCKE
Neurophysiological Determinism and Human Action: V. M. WEIL
Relative Identity: A Reply to Nicholas Griffin: H. W. NOONAN
More on Non-Epistemic Seeing: D. CLOSE
Freedom, Determinism and Character: E. SANKOWSKI
Actions are not Events: K. BACH

Critical Notice
Will, Freedom and Power and *Freewill and Responsibility*
by Anthony Kenny. T. HONDERICH

Book Reviews, Books Received, Notice

Orders to: Journals Department, Basil Blackwell Publisher Ltd.,
108 Cowley Road, Oxford OX4 1JF, England
Annual subscription:
Institutions: £10.00 or $22.50 (Canada $26.00) *Individuals:* £8.00 or $18.00 (Canada $21.00) *Single issues:* £3.00 or $6.75 (Canada $7.75)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

Erkenntnis

An International Journal of Analytic Philosophy

Editors: CARL G. HEMPEL, WOLFGANG STEGMÜLLER, and WILHELM K. ESSLER

Erkenntnis is a philosophical journal publishing papers on foundational studies and scientific methodology, covering the following areas:

- that field of philosophy associated today with the notions of 'Philosophy of Science' and 'Analytic Philosophy' (in a wide sense);
- the philosophy of language, of logic, and of mathematics;
- the foundational problems of physics and of other natural sciences;
- the foundations of normative disciplines such as ethics, philosophy of law, and of aesthetics;
- the methodology of the social sciences and the humanities;
- the history of scientific method.

Subscription Information: 1979, volume 14 (three issues). Private rate: Dfl. 45,– / US $ approx. 20.00. Institutional rate: Dfl. 125,– / US $ approx. 60.00. Prices include postage and handling.
Private subscriptions should be sent direct to the publishers.

D. Reidel Publishing Company
P.O. Box 17, 3300 AA Dordrecht, Holland
160 Old Derby St., Hingham, MA 02043, U.S.A.

Journal of Philosophical Logic

Published under the Auspices of the Association for Symbolic Logic

Editor-in-Chief: RICHMOND H. THOMASON

The *Journal of Philosophical Logic* is concerned with philosophical studies utilizing formal methods or dealing with topics in logical theory. Specifically included are: contributions to branches of logical theory directly related to philosophical concerns; contributions to philosophical discussions that utilize the machinery of formal logic; discussions of philosophical issues relating to logic and the logical structure of language; and philosophical work related to special sciences with emphasis on foundational problems.

Subscription Information: 1979, volume 8 (four issues). Private rate: Dfl. 45,– / US $ approx. 20.00. Rate for members of the *Association for Symbolic Logic:* Dfl. 38,50 / US $ approx. 17.00. Institutional rate: Dfl. 128,50 / US $ approx. 62.00. Prices include postage and handling.
Private subscriptions should be sent direct to the publishers.

D. Reidel Publishing Company
P.O. Box 17, 3300 AA Dordrecht, Holland
160 Old Derby St., Hingham, MA 02043, U.S.A.

EDITORIAL BOARD

R. B. Braithwaite
G. Buchdahl
L. J. Cohen
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
W. C. Kneale
S. Körner
J. L. Mackie
D. J. O'Connor
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow
J. H. Woodger

EDITORIAL ADVISORY PANEL

The Editors are helped by members of the Editorial Board, members of the Committee of the British Society for the Philosophy of Science and by an Advisory Panel which includes:

Peter Clark
Jon Dorling
Arthur Fine
Michael Gardner
Donald Gillies
Ian Hacking
Colin Howson
Clive Kilmister
Noretta Koertge
David Miller
Alan Musgrave
Vernon Pratt
Ian Roxburgh
Geoffrey Sampson
Peter Urbach
Elie Zahar

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1979-80

President: Mr G. Buchdahl
Vice-President: Mr L. J. Cohen
Hon. Secretary: Mr A. Young
Hon. Treasurer: Mr J. Powers
Members: Mr D. Hirschmann, Dr M. L. G. Redhead, Professor Y. Wilks, Dr P. M. Williams

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all interested persons on payment of the annual subscription. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £2.50 for U.K. residents and £1.25 for overseas residents ($2.50 if resident in U.S.A. or Canada). Those who wish to join should write to the Hon. Secretary, Mr Alistair Young, School of Social Sciences, University of Sussex, Falmer, Brighton BN1 9QH.

Brit. J. Phil. Sci. Vol. 31 No. 1, March 1980

The British Journal for the Philosophy of Science

Contents

Articles

1–43 Elie Zahar Einstein, Meyerson and the Role of Mathematics in Physical Discovery

45–62 L. Jonathan Cohen Bayesianism versus Baconianism in the Evaluation of Medical Diagnoses

Discussion

63–67 Geoffrey Sampson Popperian Language-Acquisition Undefeated

Review Article

69–79 David H. Sanford Armstrong's Theory of Universals (review of D. M. Armstrong: *Universals and Scientific Realism*)

Reviews

81–108 Irving Copi on Lewis Carroll: *Symbolic Logic*; M. J. S. Hodge on E. Manier: *The Young Darwin and His Cultural Circle*; Robert Nola on Hilary Putnam: *Meaning and the Moral Sciences*; David Papineau on T. W. Hutchison: *Knowledge and Ignorance in Economics*; D. C. Phillips on Richard Mattessich: *Instrumental Reasoning and Systems Methodology*; Ronald Rubin on B. Williams: *Descartes: The Project of Pure Enquiry* and E. Curley: *Descartes Against the Skeptics*

Notes to Contributors

8 SEP 1980

Printed in Great Britain at Aberdeen University Press

Volume 31 Number 2 June 1980
BJPIA5 30 (No. 1) 111-212 (1980) ISSN 0007-0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: JOHN WORRALL

Subscriptions

The *Journal* is published quarterly. The subscription price for mailing addresses in Britain is £7.00 net per volume post free and for mailing addresses outside Britain £11.00 (or foreign currency equivalent calculated at the day's rate); single issues £2.50 net (overseas £4.00) plus postage. Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

For members of the British Society for the Philosophy of Science—who should order direct from Aberdeen University Press—the subscription price is £4.50 (overseas £7.00). Membership of the Society is international and is by no means confined to people resident in the United Kingdom.

Microform

The current volume of the *Journal* is available in microform to subscribers to the printed version, *at a combined print/microfiche rate of* £12.50 (*overseas* £16.50).

Back Volumes

Volumes 1–16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17–30 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application. Volumes 1–29 are also available in microform from Kraus Reprint Corporation.

Typescripts and Review Books

Typescripts, correspondence and books for review should be sent to the Editor at the Department of Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street, London WC2A 2AE.

Review books are accepted on the understanding that, although they will appear in the 'Recent Publications' list, a full review cannot be guaranteed, and that books not reviewed are not returnable.

Notes to Contributors

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Detailed instructions on the lay-out of typescripts are published occasionally at the end of the *Journal* and anyway may be obtained from the Editor.

Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

© British Society for the Philosophy of Science 1980

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editor.

Brit. J. Phil. Sci. 31 (1980), 111–129 *Printed in Great Britain*

Brain States

by KATHLEEN V. WILKES

1 Philosophers working on the mind-body problem have always interested themselves primarily in the relations that hold, or fail to hold, between the mental and the physical. Their theories are, indeed, identified in terms of the relation for which they argue; so we find identity theories (of several different varieties), interactionism, eliminative materialism, functionalism, parallelism, epiphenomenalism—and the list continues. Strangely enough a prior question is usually ignored, namely just what are these mental and physical relata, the items that provide the two terms of the relation? It is true that the mental relata tend to receive *some* notice; for instance, 'the mental' is often divided into two broad categories—the category of the occurrent, incorrigibly-known sensation-type phenomena, like pains, on the one hand, and on the other the category of the intensional propositional attitudes, such as beliefs and desires; but details of the membership of either of these categories are rarely provided, and discussion of the mental relata stops there. The physical relata are hardly discussed at all. This latter neglect, the neglect of the physical relata, is presumably due to a laudable enough desire not to meddle with scientific matters lying beyond the philosophical horizon. However, there is a body of fact and theory, readily accessible to the layman, that deserves and repays attention; it can shed a fascinating light upon the whole mind-body debate. (*E.g.*, Dimond [1972], Gardner [1977], Geschwind [1974], Hinshelwood [1900], Luria [1973], Valenstein [1973]).

In this paper I discuss some of the implications for the mind-body problem of recent work in neuropsychology. There is one essential preliminary caveat: the argument touches upon only such theses as concern themselves with the possibility of finding *lawlike* correlations between mental states and physical states. I am assuming that these are the most interesting theses, using 'interesting' in a sense that betrays a bias in favour of the scientific concern with the mind-body problem; for the question of the viability of lawlike correlations between mental and physical lies at the heart of future research into the psychophysiology of human experience and action, and is presupposed by any discussion of the reducibility[1] in principle or practice of mental states to physical

---

[1] Any interpretation of 'reduce'—and there are very many in the literature—will fit the argument.

Received 26 *February* 1979

processes in the brain and central nervous system. Hence, for example, Davidson's 'anomalous monism' (Davidson [1970]) has a positive thesis that is not interesting in this sense, for it argues for only particular identities between mental and physical, rather than for lawgoverned general links; but it has an interesting negative thesis, in that it claims that there can in principle *be* no genuine laws between the two sets of relata. My concern, then, is with lawlike correlations, and with how certain facts about brain states affect not only their form and content but their very viability.

2 If it is agreed that a scientifically interesting theory about the relation of mind and body presupposes that there can be lawgoverned links forged between the two, or else proved to be impossible in principle to forge, then instantly there appear some obvious but nonetheless powerful constraints upon the relata. The first constraint is that the two domains—of mental and neurophysiological phenomena—must each comprise clear and explicit taxonomies of items, such that within each domain the conditions for their identification are apparent and effective, or are thought to be capable of becoming so.[1] (For example, the theoretical concept of an 'engram' is at present vastly obscure, but can still be used so long as it is believed that one of the competing theories of learning will manage to specify it precisely within the terms and postulates of the theory, declaring it to be a matter of changes in the structure of the cell membrane, *or* of functional changes in synaptic chemistry, *or* of structural changes in intracellular macromolecules, *or* of some combinatory compromise between these three theoretical approaches (see Dunham [1977], pp. 158—79). The notion of an engram would be worse than useless, indeed positively misleading, unless scientists thought that it held the promise of ultimate precise specification.) Second, the domains in question must be worthwhile domains. By 'worthwhile' in this connection I mean that they must be such as to contribute to the theory of which they form the ontology (thus, for example, neurons with axons of eactly 8⅝ mm long, or the bumps of the phrenologists, would presumably not belong to a worthwhile domain); the mental items and properties must be those in terms of which human behaviour and experience can be satisfactorily described and explained, and the neurophysiological items and properties must be

---

[1] As will become clear later on (section 4) the creation of these taxonomies in each of the two sciences is and must be heavily dependent upon the work in progress in the other science; for example, psychologists hope to postulate processes that are capable of neurophysiological realisation, and neurophysiologists try to identify just those cerebral events and structures that meet the psychologists' demands.

capable of providing the same service for the understanding of central nervous system functioning. In short, the conceptual apparatus of each investigation should be the best (or one of the best, since the profusion of both psychological and neurophysiological theories shows that at least at present there is gross underdetermination of theory by data) for the descriptive and explanatory purposes of the science.

More could be said about the constraints that any lawgoverned relationship must impose upon its relata, but these are indisputable and will suffice for present purposes. It is worth investigating how they affect one fairly prevalent bunch of theories (less prevalent, perhaps, than it used to be) about the mind-body relation: that is, that individual mental states, such as the thought that one must buy more toothpaste, or the recognition that the object on the horizon is a windmill, may be correlated—whether through identity, parallelism, interactionism, functionalism, or what you will—in a lawlike fashion with cerebral states. I shall discuss presently whether such states as these are suitable for membership of the ***mental*** ontology; we shall first consider the nature of their neurophysiological correlates.

We are searching for a description of a brain state (or, more leniently and plausibly, of brain states; there is nothing to prevent lawlike relations between mental and physical being one:many rather than one:one) such that the cerebral processes in question may be identified and reidentified, such that the description picks out brain states that would be worthwhile members of the neurophysiological ontology, and such that we have reason to believe that one of the set of correlated states will recur whenever anyone remembers his need for toothpaste or recognizes a windmill. What seems absolutely clear from the scientific evidence is that there could be no such states. To see this, consider how just one state of the necessary kind might be identified. It is now generally acknowledged that any mental phenomenon (whether it be a cognitive state, a simple sensation, an emotion, or feelings of hunger and thirst) will be embodied by or grounded in[1] a variety of processes in distinct and dissimilar areas of the neocortex and/or limbic system. In each such area the neurons may be of two or more different kinds, and will be capable of firing at varying speeds and in different combinatory patterns. If we were to take just one of these brain areas, the activity of just one type of neuron in it, and somehow contrived to sort out the neural activity relevant to our interests from the background 'noise' of unrelated neural discharge, then we could attempt to specify this small locus of cerebral activity by (let us suppose) observing neurons *a*, *b*, *c*, . . ., *n* to be active, neurons *x*, *y*, *z*, . . ., *m* to be inhibited, the speed and the pattern of neural firing to be thus-and-so (synchronous, sequential, waltz-time, *etc.*), and the duration of the activity to be *D*. To achieve this description we would need, of course, to do more than merely examine the facts of neural firing. For to discover whether neuron *x* is inhibited or simply inactive we should probably have to go into the question of how much of which neurotransmitter-substance was being released in what areas, and maybe would have to go into yet more details of cellular chemistry. However, let us suppose it done. Then we would repeat the process with the other sorts of neuron in the same area (if any); then would add that completed description to similarly complete descriptions of the activity in the other relevantly-active brain areas and of the activity in the connecting neural links between them. The finished description would pick out 'the' brain state that correlates with the thought that one must buy more toothpaste, or that the object on the horizon is a windmill.

---

[1] 'Embodied by or grounded in' is left deliberately vague, so that as many as possible of the rival monistic theories may be covered. Dualists are free to substitute 'regularly correlated with'.

It is patently obvious that such identity-conditions for brain states are wholly unrealistic in their stringency. The description-schema outlined would pick out a brain state that would never, and probably could never, recur, and hence a brain state that could not be a worthwhile member of the ontology of a science of neurophysiology; there are a number of reasons for this. First, many of the individual neurons listed may have decayed—they do not replace themselves, and after one's late teens several hundred decay each day. Second, the neurons involved may not be free the next time that the individual thinks that he must buy more toothpaste. Neurons are not 'toothpaste specific'; they do not lie idle in the interim, waiting until toothpaste is again considered. Most of the cells involved in the activity outlined above will be activated again and again, in different patterns and combinations, for different cerebral functions; so when the subject thinks a second time that he must buy more toothpaste, an enormous number of the neurons involved in his first thought will be otherwise engaged, or inactive in their refractory rest period. Third, the micro-organisation of the brain changes with time and experience. A pattern of neural firing may be triggered off, repeated, and established; but it is all the time liable to mutation, and will be modified by new but related inputs, or may become 'coded' for reasons of efficiency into a more economical pattern (it is unlikely, for example, that the visual cortex of the ordinary toothpaste-buyer will need to generate the cerebral correlate of a mental image of a tube of toothpaste as he remembers his need for it; and even if it does,[1] such an image may be more or less schematic on any given occasion). Every new input modifies, even if only fractionally, the neural organisation upon which it impinges. There is good evidence for this from temporal lobe stimulation. Such stimulation may evoke the same memory or hallucination on more than one occasion, but usually only if there is a very short time (up to a minute or two) between stimulations; if a longer time elapses between stimulations of the same area by the same (unmoved) electrode, a new memory or hallucination tends to be reported (see Valenstein [1973], pp. 110–14). Nothing is less static than the nervous system.

That there should be cross-brain identity of brain states picked out by the above description-schema is, evidently, even more implausible. Not only do brains differ considerably in size and hence, presumably, in the total number of cells they contain (Anatole France's brain weighed 1,100 grams, half the weight of Byron's at 2,200 (Sagan [1977], p. 35 f.)), but they differ also in functional organisation. There are conclusive proofs of this. Left hemisphere dominance (by which I mean dominance for certain functions only, primarily linguistic and analytic ones; the right hemisphere is dominant for others) is totally absent in about a tenth of the population, and the degree of its dominance varies widely in the remaining nine-tenths. Since linguistic processes seem to play a large part in a great many cognitive abilities (see Bruner, Goodnow, and Austin [1956], *passim*; Luria [1973], chs. 12 and 13), the variation in the degree of cerebral dominance will carry with it a corresponding variability in the organisation of cognitive processes from person to person. The evidence of electrical and chemical stimulation of the unanaesthetised brain, and the evidence of lesions to the brain, strongly suggest the same conclusion. For example: stimulation of the septal region of the brain may in some people provoke pleasurable sexual sensations, but in others can arouse terror or anxiety (Valenstein [1973], p. 105 f.); no reliable correlations can be drawn between stimulation to the amygdala and the behavioural responses (see Heath, Monroe, and Mickle [1955]); patients with very similar lesions to the same areas of the prefrontal lobes may become either hyperactive or apathetic, there is no predicting which (see Teuber [1972]); there is a marked disagreement on whether psychosurgical operations on the anterior cingulum of the brain relieve extreme anxiety and depression or not (contrast Foltz and White [1962] with Scoville [1971]). If there is so much difficulty in correlating broad and unspecific phenomena, such as moods, with stimulation of or damage to areas of the brain, it will clearly be yet more impracticable to hope to correlate cerebral processes with fine-grained events like thoughts about toothpaste. Even more important: the *micro*-functioning of brains will be as unique to the individual as are his fingerprints. Every stimulus has some effect on the brain, modifying, reinforcing, or disrupting both the existing neural patterns and the laying-down of new patterns—and evidently no two people have the same histories of sensory input. How each input is coded will be determined by a great number of factors, including the past history and neural patterning of the relevant cerebral areas and by the familiarity of the input (the brain finds 'short-cut' codings for recurrent, and hence familiar, inputs). In sum, brain states such as those suggested could play no role at all in the study of the organisation of the brain; hence they will not belong to the ontology of neurophysiology. There can be no worthwhile correlations between specific thoughts and brain states.

---

[1] There is some evidence that the visual cortex is involved in many more 'purely cognitive' activities than one might at first suppose; it perhaps does form the cerebral correlate of a mental image of a tube of toothpaste when we think about toothpaste, even if we are aware of no such image. See Luria [1973], pp. 315 ff., and Posner [1969].

This is, in fact, a conclusion that could have been reached without benefit of science, simply by considering what kind of *mental* ontology any systematic study of human experience and behaviour would require. Specific thoughts could have no part in this. Thoughts are intensional, and are therefore difficult to identify or reidentify—we do not know whether Joe's thought about Cicero is a thought about Tully or not, still less whether it is the same thought as Jim's thought about Tully. Moreover, since the mental realm is holistic, to attribute any specific thought to an agent presupposes that one is already tacitly ascribing to him a host of other thoughts (beliefs, desires, *etc.*), in terms of which the explicitly ascribed thought makes sense. This means that *what* thought we ascribe to anyone will depend as much, if not more, on what we already know and do not know about him—our implicit theory of his behaviour—as upon what thoughts, if any, are currently and consciously passing through his mind. In short, the mental realm, the ontology expressed in ordinary language, is not an ontology that purports to explain human experience and behaviour in any systematic way. It is admirably tailored to fit our non-scientific demand to predict or explain what a given individual, about whom we already know a certain amount, may do at a particular time and in a given set of circumstances, but is of little help in accounting for the behaviour of human beings in general. For that we need science.

Many will think that I have spent too long flogging an already moribund horse; several philosophers have pointed out that the enterprise of correlating individual propositional attitudes with states of the brain is implausible in the extreme (*e.g.* Nagel [1965], part iv; Taylor (1967]).

But the time spent has not been wasted. It has told us something about what brain states are not; and has incidentally shown that the hypothesis of 'brain reading' (the spy with the cerebroscope) is not a threatening one. If God could look into our brains he would not know what we were thinking. We can now pass from the negative thesis to a more positive one, and consider just what God—or the cerebroscopist—might be able to discover from a study of cerebral activity.

3 To see what brain states might be, it will be profitable to describe in outline some of the various brain processes that correlate with one kind of mental activity, the activity of seeing. Perception is a good example to take, as scientists have recently made great strides in understanding much of the process, even though a good deal remains dark. It is, for instance, much better known than are the processes of auditory perception. Moreover perceiving, although it undoubtedly has its 'purely sensory' side, touches upon cognitive abilities in so far as perceptual *recognition* is essential for full seeing, so that a study of perceiving cannot fail to shed some light upon the mysteries of thinking.

The optic nerves travelling from each eye terminate (with one relay in the lateral geniculate nucleus of each hemisphere) in the primary areas of the occipital cortex, right at the back of the brain. Fibres from those cells that have received input from the right half of each eye's visual field end up in the occipital cortex of the left hemisphere, while those dealing with the left half-field end up in that of the right; so approximately half the fibres from each eye have to cross over each other on their way to the occipital lobes. When they reach the primary visual cortex the fibres fan out (the optic radiation) in such a way that the visual field is projected in a strictly somatotopical manner: the upper half of the field is projected at the base of the optic radiation, and the lower half at the top—thus, simply, the optic radiation of the left hemisphere encodes the right half of the visual field upside-down. The neurons of the primary visual cortex are highly specialised. Practically all of them are specific to vision; nearly all of them, like the retinal cells, react only to determinate features of the input—some reacting to colour or shade contrast, others only to convex or concave edges, others only to bar shapes, others only to bars on a particular alignment, and so on. This first stage of the visual process, therefore, preserves the discrete and detailed cues into which the retinal cells have broken-up the visual field, and locates those cues in one:one correspondence with their place in the visual field. Hence, as we would expect, stimulation of cells in this area evokes simple hallucinations—of spots, flashes, lines, or flames (Luria [1973], p. 114). Localised damage causes blind spots in the visual field; so long as the lesion is not extensive the rapid alteration of eye movements in the saccadic jumps can usually (although not invariably[1]) compensate for these spots so effectively that the individual may not even realise that he has them. Of this area, then, we can say that God, or the cerebroscopist, once they had worked out which neurons were specific to which sorts of input-features, could tell from the patterns of neural activity what the raw data of the perceptual field were,

Above and overlapping the primary occipital zones lie the secondary zones, deeper in the brain, which receive input from the primary zones.[2] Neurons in this area are still almost exclusively specific to sight, but are much less specific than are the cells of the primary cortex to features. They have shorter axons and are known as 'associative neurons', for associating is precisely their function: they arrange and synthesise the raw elements of perception transmitted from the primary visual cortex into complex and organised wholes. Good evidence for this comes from experiments upon commissurotomy patients; when presented with a stimulus (for example a photograph of a face) of which only one half has been shown, these patients will say unhesitatingly that they have seen the whole object (Trevarthen [1974], p. 254 f.); perception always 'goes beyond the information given'. That it is the secondary visual cortex that is primarily responsible for composing these Gestalts is supported, albeit indirectly, by the findings from both stimulation to, and lesions of, this cortical area. If the secondary visual cortex is stimulated the subject usually experiences *complex* visual hallucinations—of people, or of animals, trees, or buildings—in other words, the stimulation evokes patterns of neural activity that are fairly well-established patterns in that individual's history, whole complex patterns that get triggered off by an initial stimulation of a small subset of the cells involved. Lesions, on the other hand, can destroy the ability to see, draw, or picture whole complex objects. Luria tells of a patient who could look at a picture of a pair of spectacles and note the individual components (two circles, a cross-bar, projecting flanges), but who then asserted that it was a picture of a bicycle; and of another who saw a telephone with its dial but identified it as a clock (Luria [1973],

---

[1] Some lesions to this area in the right hemisphere go unregarded and there is no compensatory activity; see Brain [1941].

[2] One should insert a caveat here that will apply both to the distinction now drawn between primary and secondary visual areas and to that shortly to be drawn between the secondary areas and the parieto-occipital region. The boundaries—whether of structure or of function—are neither sharp of themselves nor, predictably, do all neuropsychologists agree on where they should lie (see Teuber [1955], especially p. 278). I am in general following Luria's [1973] description of the functional organisation of the visual system.

p. 116 f.). When these patients attempted to draw a complex object, they tended to draw a 'visual list' of the components—so that, for example, a patient might depict the four legs, head, body, trunk, and tail of an elephant, but in no sensible relation one with another; and the capacity to visualise, unless the visualised object is extremely simple, vanishes entirely.

A further interesting feature of these secondary zones is that the two hemispheres begin to divide the labour. The left hemisphere (*LH*), as is well known, is in most people dominant for speech and all language-related abilities[1]; the secondary visual cortex of the *LH* seems, therefore, to be largely in charge of the recognition of letters and words. A lesion here, even though it may leave unscathed the ability to perceive Gestalts, can disrupt the recognition of letters while leaving untouched the recognition of objects; lesions to the right hemisphere (*RH*) very rarely have this effect. On the other hand a lesion to the secondary zone of the *RH*, while also leaving unimpaired the perception of complexes, may produce the fascinating phenomenon of prosopagnosia—the inability to recognise faces. The patient knows that a face *is* a face; but whose face it is, even if it is the face of a friend or relative, or his own in a mirror, he is unable to say (Bornstein [1963]; Gardner [1977], pp. 153–6). Other lesions to this area of the *RH* may disrupt the recognition of familiarity, when patients fail to recognise some object as their own—although once they handle it, it is soon enough recognised. Lesions to the corresponding part of the *LH* do not have this effect.

There seems to be little that God or the cerebroscopist could discover from a study of activity in the secondary visual cortex. It is the unique micro-organisation of the individual brain that determines just what neural activity in the secondary visual cortex correlates with the synthesising of circles and lines into a pair of spectacles; it is the unique history of the individual's experience that determines what neural activity underlies the recognition of familiar people or objects. All that could be inferred from the study of micro-processes in the secondary visual areas would be that *some* fairly complex perceiving (or visualising or hallucinating) was going on; but no more.

Full-blooded perception requires yet more than this; but for part of the extra we have to turn to the highly complex and still highly mysterious area of the angular gyrus. This area borders on, and overlaps, the occipital, parietal, and temporal regions, and so its visual functions mesh with other sensory abilities and with linguistic processes. The nearer we are to the parieto-*occipital* part of the angular gyrus area the more, generally speaking, we will find complex but still primarily or distinctively *visual* functions; but even then they tend to be infused by contributions from other sensory modalities, and by the language systems.

---

[1] To say that it is dominant for linguistic functions is neither to assert that the *RH* has negligible linguistic capacities nor that it is not heavily involved in language processing. The linguistic capacities of the *RH* are explored in Hecaen [1978] and Zaidel [1978]; that it is as active as the *LH* during linguistic processing is indicated by measuring the blood-flow by radioactive isotopes; see Lassen, Ingvar and Skinhøj [1978].

For example, it would seem that the parieto-occipital area is largely responsible for the ability to organise perceived objects in visual space, so that the position of objects in relation to one another on the frontal, sagittal, and horizontal planes can be accurately detected. Moreover, it seems to be required for 'simultaneous synthesis'—an activity whereby a whole situation, of which the individual items have been processed serially by the occipital cortex, may be perceived as a single whole at a single time. Evidence for these conjectures again comes from patients with damage to this area; such people are often unable to tell the time from the position of the hands on a clock, cannot say which elements of a visual array are above, below, behind, or to the left or right of, other elements; they fail to see a *group* of objects, and are frequently unable to work out the theme of a moderately complex narrative picture (Luria [1973], p. 234). These same lesions disrupt not only the visual functions mentioned, but also interfere with a host of other abilities that we do not tend to regard as particularly 'visual'. If the lesion is to the dominant hemisphere, for example, then not only spatial relations but also logical, grammatical, mathematical, and temporal ones may be hard to grasp. Sentences containing words like 'before', 'under', 'therefore', 'afterwards', or incorporating relative or propositional clauses, passive verb forms, or the possessive (*e.g.* 'father's brother') are rarely understood; and operations of addition and subtraction are performed badly (Luria [1973], pp. 151–6). Although, then, we can legitimately say that this area is involved in complex visual processing, it may prove more perspicuous to describe the disrupted ability as the recognition of relations in general—an ability that includes *inter alia* visuo-spatial relations.

Another *LH* function where vision merges with non-visual capacities is reading. The clearest evidence for this comes from patients with a rather specific lesion to the parieto-occipital zone of the *LH*.[1] Such patients are those with pure alexia—a condition that disturbs the recognition of words, or of letters, or both; sometimes with, but often without, preventing the recognition of objects and people, of printed numbers (Hinshelwood [1900], p. 12), or (perhaps) of musical notation.[1] The only other capacity that is regularly disturbed along with reading is the capacity to name colours on sight (Geshwind [1974], ch. IX). It is difficult to know whether such difficulties should be called 'visual' or not. The pure alexic clearly *sees* the symbols '*DIX*' when he reads them as '509' rather than as the syllable 'diks' (Gardner [1977], p. 16); besides, he can copy letters (although, since he treats them as shapes rather than as letters, the product is less fluent than is his spontaneous writing (Geschwind [1974], p. 395 f.); and some patients are alexic for letters but not words, or words but not letters (Hinshelwood [1900], *passim*), or English but not French words (Gardner [1977], p. 121), all of which augurs normal visual acuity. Even if he cannot name colours on sight, he remains able to match and sort them accurately enough (unless he has an independent defect in colour vision (Geschwind [1974], ch. IX). So, in a sense there is 'nothing wrong with' his sight; however, in as much as something very odd is going on with the processing of verbal materials of a certain type,[2] the phenomenon should be included in our account.

---

[1] The lesion is one that destroys the visual cortex of the *LH* and penetrates deep into the white matter of the parieto-occipital junction, thus cutting off the visual data of the visual cortex of the *RH* from the language centres of the *LH* (Geschwind [1974], p. 384). In fact the parieto-occipital area can be spared and the same pure alexia produced by a lesion that damages the splenium (the lower end of the corpus callosum), for this too cuts off the *RH*'s visual cortex from the *LH* (Geschwind [1974], pp. 4–15).

The contribution of the parieto-occipital zone of the *RH* is even less well understood; it may indeed be that its functional organisation is not differentiated to the same extent as is that of the *LH* (Luria [1973], p. 164). It seems safe to say that it plays a *larger* part than the corresponding area of the *LH* in many visuo-spatial activities: finding one's way around a familiar locale or building, detecting line-orientation and perceiving depth —in general, organising three-dimensional space—and there is some evidence that it is required for the individual to relate visual stimuli to his own past experience: a patient of Luria's with a right-sided lesion, when looking at a photograph of a group of soldiers sitting on a tank, described it as a picture of her father, sisters, and children (Luria [1973], p. 126 f.). But the scope and powers of the *RH* are as yet largely unknown.

The last important cerebral area to mention in a discussion of the cerebral basis of perception is the region of the frontal lobes. It is impossible to exaggerate the range and diversity of the functions of these lobes; but one of their many tasks is to control the *activity* of perceiving, stamping upon it its directed and purposive character. This area, by controlling the saccadic eye movements, ensures that a complex visual object is scanned efficiently and economically, and is actively searched for clues to its identity; it forms hypotheses[1] about what the object might be (so, to go back to the patient mentioned earlier with damage to the secondary visual cortex who guessed that a picture depicted a bicycle rather than a pair of spectacles, he was at least able to form an intelligent conjecture based on inference from what he saw, as his frontal lobes were undamaged) and in terms of these hypotheses it guides the searching movements; in short, it makes perception a goal-directed activity. Thus, lesions to the relevant parts of the frontal lobes will not disturb *simple* perception; the patient is probably able to perceive and recognise elementary pictures, to read short and easy sentences, to read numbers, and to recognise faces and familiar objects. But difficult visual objects, that require some interpretation, might not be recognized or understood. A hat upside-down might be seen as a plate; an ambiguous picture, like that of the duck-rabbit, may be seen either as a duck or as a rabbit, but once it has been seen as one cannot be seen as the other; complex narrative pictures are not understood; and it often proves difficult if not impossible to perform such actions as tracing with a finger the outline of a cross on the black and white squares of a chessboard (Luria [1973], pp. 240–2). Experiments have shown that the eye movements of these patients, when they are studying a picture, do not display the controlled and economical searching movements of the normal perceiver but instead wander about at random, failing to pick out or concentrate upon the significant features of the display (Luria [1973], pp. 216–18). Were it not for the frontal lobes, it is clear that perception would be a vastly impoverished matter.

---

[1] See Hinshelwood [1900], p. 20. Geschwind, however, finds alexia without amusia to be a rare and probably non-existent condition (Geschwind [1974], pp. 396–7); and it has to be admitted that Hinshelwood cites no cases of it.

[2] I examine this oddity in 'Brain Lesions' (forthcoming).

We have mentioned four distinguishable areas of the cortex that are involved in perception (four if the corresponding areas of *LH* and *RH* are counted as one), and have touched upon a few of the very diverse activities in which the brain must engage if its owner is to see and recognise items of heterogeneous types. Using this as a model, we can go on to consider what descriptions of what brain states might be expected to correlate in lawlike fashion with what descriptions of what mental states.

4 It is evident at once that we can rule out the notion that any single brain area is 'the' locus for any mental state, event, or process. Localisation had its heyday in the early nineteenth century, when Gall and Spurzheim drew their maps of the brain, in which they would allocate one clearly defined and discrete region to 'amativeness', others to 'veneration', 'mirthfulness', 'firmness', 'secretiveness', 'self-esteem', 'hope', and even

---

[1] The use of this apparently anthropomorphic description will be defended shortly.

'marvellousness' (Gall and Spurzheim [1810–18]). Neuropsychologists can indeed draw what may well be described as 'functional maps' of the brain, filled out in greater or lesser detail; but in such a map what will correspond to 'perceiving' will be a complex series of processes, involving at least four brain masses that are functionally and topologically distinguishable, a series within which each contributory structure has its own essential roles. Even the cerebral correlates of the mental events that seem intuitively most simple—*i.e.* minimally cognitive—such as the experiencing of hunger, thirst, and pain,[1] cannot begin to be localised in just one area.

The next obvious point to make is that the battery of ordinary language mental terms cannot all be correlated with distinct series of cerebral processes, no matter how complicated such processes may be. This should be evident; nothing in the brain can distinguish seeing from looking at, knowing from believing, wanting from desiring, intending from purposing, remembering from recalling, and the like; such distinctions, which the richness of ordinary language mental terms allows, could not be captured in neurophysiological terms. It is easy to see why this must be so: the diversity and wealth of ordinary language provide not only for the purposes that the vernacular shares with science (description and explanation), but also provide for countless *non*-scientific purposes. We say, for example, '*X* knows that *p*' if we want to endorse his belief, but may choose '*X* thinks that *p*' if we are sneering at him; by means of a careful selection of terms we can applaud, condemn, blame, correct, deprecate, urge, compliment, condole, plead, grade, and do many other things far too numerous to list. The choice of terms is just as much a matter of our attitude to the agent as it is a matter of what (if anything) we conjecture to be passing through his mind at the time. Any mental state that we might realistically expect to relate in lawlike fashion with a series of processes in the brain must be distinct and distinguishable from any other such state; hence the vast majority of ordinary language mental terms will not enter into the laws relating mental to physical except indirectly—except, that is, insofar as they can be understood in terms of the concepts that *are* to be related to neurophysiological descriptions.

A third obvious point, related to the last, is that the members of the privileged set of mental concepts that remain as *explananda* must be freed of many of the conceptual links with which vernacular usage endows them. For example, 'see' is usually taken to be an achievement ('got-it') verb, and hence, it is sometimes said, cannot be a time-taking *process*; and many would claim that it is inapplicable unless there is a public object that is seen. Similarly 'remember', as strictly used, implies that what is remembered is indeed the case. One could cite many more of these common implications. Evidently, though, such linguistic conventions will not do for the purposes of psychology or neuropsychology. For their purposes, seeing has to be considered to be a complex and highly intricate constellation of processes, a constellation that is active with only minor variations whether what is 'seen' is a public object, an illusion, a dream, an image, or a hallucination; while remembering will correlate with the same sorts of processes whether what is remembered did indeed happen or not.

---

[1] The philosopher reluctant to plough through the scientific literature on pain will find its complex structure described in Dennett [1978].

The next point, again closely related to the last, is that (neuro)psychology needs a vocabulary to describe more *kinds* of entity than the mental concepts of the vernacular need to do. For ordinary language mental concepts are, almost by definition, ascribable only to whole sentient organisms; whereas psychologists and neuropsychologists, as we have seen, need to describe the capacities and operations of various areas of the brain as well as the behaviour of the individual person. Crudely, it is necessary to describe what the angular gyrus region *does* when one is analysing the capacity of the person to see. What psychology does when confronted with this problem is exactly what all other sciences do; it borrows terms from the vernacular, but having adopted them, will then adapt them to fit its descriptive and explanatory tasks—freeing them from overlap with other concepts, freeing them from the network of semantic implication that everyday usage has bestowed upon them, and trying to delimit their new sphere of operations more explicitly than is common or necessary in our day-to-day vocabulary. (Compare the way that physics has handled 'force', 'mass', and 'energy'.) One of the implications that this process of denaturalisation may have for at least some of the borrowed concepts is that their subject need no longer be a whole organism; one can unashamedly describe the operations of, say, the parieto-occipital area by means of terms which, *qua* ordinary language terms, are misused or metaphorical, but *qua* scientific terms are legitimately literal (in a well-developed conceptual structure[1]). Hence, brain structures may be said to 'code' information *without* the scare-quotes; or may be characterised as analysing, interpreting, synthesising, grouping together, checking, correcting, scanning, and so forth, without criticism.[2] (It goes without saying, of course, that the more complicated the functions ascribed to complicated brain structures, the more it is clear that we need still to do; a complex function performed by a complex structure is something which in its turn demands analysis and dissection into less and less intricate functions performed by less and less intricate structures. It is a sign that we are getting close to the nitty-gritty neural level when we can talk in terms of the functions of identifiable groups of cells—the primary visual cortex will presumably be the first of the cortical areas involved in vision to permit such understanding of its microstructure.) It is vital, then, to note the distinction between 'the mental' (the vernacular) and 'the psychological' (the scientific); failure to do so is to forget that psychology *is* a science.

---

[1] There is, of course, no such consistent and generally-accepted conceptual system at present. I am talking in a hortatory rather than a descriptive manner about the conceptual structure at which neuropsychologists *should* aim.

[2] I have discussed this point elsewhere in greater detail (Wilkes [1974]). A similar line is taken by D. C. Dennett (Dennett [1978]), and, more explicitly, in Dennett [1975].

These four considerations suffice for the first conclusion: that the 'mind' relata of the mind-body problem are not the events, states, and processes picked out by ordinary language mental terms, but are rather those events, states, and processes that are discovered or postulated by the science of psychology, and are articulated in the conceptual apparatus of that science. Terms to describe the behaviour of an individual organism at the global level have to be denaturalised just as much as the terms adapted for the purpose of describing the functions of, say, the frontal lobes; seeing, writing, reading, calculating *etc.* all have to be stripped of more or less of the baggage of grammatical-conceptual implication (baggage that is essential, of course, for their everyday roles) if they are to be adequate as concepts that (*a*) can describe and explain human behaviour perspicuously and systematically—that is, scientifically; and (*b*) are tolerant of their dissection and analysis into the contributory functions that mesh to realise them. Their use in psychology seems at first so close to the ordinary language use—the layman and the psychologist will almost always agree on whether or not someone is seeing or reading something—that the real and important distinctions get overlooked; but the mental and the psychological differ substantially (as is only to be expected when one reflects upon their very distinct functions) and the difference emerges clearly enough on any close examination.

If this is true, then a second conclusion follows, giving a completely different twist to the question of whether the mind can be related in lawlike fashion to the body: it is no longer, for scientists, an open question at all. For we no longer have, in advance, all the items on one side of the disputed relation, the 'mind' side; we have at best a very few of them—'seeing', 'hearing', 'remembering', 'experiencing pain', 'hungering', 'thirsting', and a few more. As for the rest of the 'mind' relata, we discover just what they will have to be (and the 'body' relata as well) *as a result of* studying the links that can be shown to hold between psychological and neurophysiological events, states and processes. Put another way, the development of both psychological and neurophysiological theories will be bound by two constraints: first, that they suffice to explain the facts of human and cerebral behaviour, respectively; and second, that they can be connected in lawlike fashion with one another. The reason for this second constraint is *not* however mere optimism that psychological theory can indeed be grounded in neurophysiological theory. It is rather that psychological theory needs neurophysiological theory as much as, if not more than, neurophysiology needs psychology. The discussion of the cerebral basis of perception bears out this contention. Electrical and chemical stimulation of the brain, the suppression by sodium amytal injections of the functioning of some areas, and the study of brain damage, have produced evidence enabling us to discover what the functions are—the *psychological* functions—that various structures of the brain fulfil, and by fulfilling which they enable the person to see and recognise a complex object; findings like these unfold much of the hitherto concealed psychological structure of concepts such as 'seeing', 'calculating', 'reading', 'hearing', 'remembering', or 'experiencing pain'. We have seen how the neuropsychological data have enabled scientists to conjecture that perception involves, *inter alia*, that parts of the brain analyse the visual input into the raw data of edges, colour contrasts, concavity, and so forth, that other regions synthesise this into an organised and structured whole, while yet other parts position it in a visuo-spatial array and classify and code it, often under the categories of language; while still another region guides the whole process by purpose and conjecture. Psychology alone might have guessed at this structure, but there could be several rival theories other than the conjectured one that sufficed to explain the observed behaviour; neuropsychological data can often cut down the number of competing theories to a manageable minimum, and hence reveal the concealed psychological structure of these global processes more effectively than can sciences that do not go inside the brain. Neuropsychology has revealed some unexpected things about the structure of psychological processes, facts that would be virtually unobtainable without it: for instance, that the reading of words has a different structure from the reading of numbers; that linguistic competence is probably much more heavily involved in recognition than one might have guessed; that the hearing of melodic tones requires a rather different analysis from the hearing of non-melodic materials; and that apparently distinct abilities, such as reading a clock, telling left from right, identifying one's fingers, grasping grammatical and logical relations, finding one's way around a building, and arithmetical calculation, must share a largely similar structure. Conversely, where neuropsychological data remain as yet inadequate or inconclusive—as seems to be the case with the study of memory—there psychological theory runs riot and is notable chiefly for an abundance of conflicting and often implausible hypotheses; a recent treatise on human memory postulates no fewer than twenty-five types of memory storage (Dunham [1977], p. 142 f.), and it would probably be possible to find psychologists who were prepared to argue for each of the options one to twenty-four. In short, psychology relies upon neurophysiology whenever it goes into secure psychological detail; small wonder, then, that these two sciences beg the question of the mind-body problem in advance. Indeed, the observant reader may have noticed that I have often used 'neuropsychology' where one might have sometimes expected 'psychology' and sometimes 'neurophysiology'. This was deliberate, and encapsulates the dissolution of the mind-body problem; no longer do we have *two* sciences which must somehow be forced to fit each other, but instead a single study, neuropsychology (or 'psychophysiology') whose molar level we can if we wish term 'psychology', and whose micro level we can call 'neurophysiology', and which straddles comfortably what was thought to be the no-man's-land of the intermediate region. Itself being neither psychology nor neurophysiology but overlapping both, its concepts, similarly, are neither; so it tolerates no psychological-physical *dichotomy*.

The final conclusion to draw from this discussion of brain states has been mentioned in passing already, but it may as well be explicitly restated now. It is the point that *individual* psychological occurrences, whether or not classified into types by the ordinary language criterion of 'same content' cannot and should not be thought of as entering into lawlike relations with brain processes. Indeed, the point is obvious; but few philosophers (Aristotle excepted) seem to have acted on it. It is, of course, the psychological *capacities* of human beings with which psychology is concerned, those capacities by means of which human experience and behaviour can most readily be explained. Similarly, the medical researcher is not primarily concerned with how and why Fred 'caught a cold' at time *t*; he wants to know how best to classify the illnesses of which the vernacular term 'a cold' is the common lay description, and, having done that, seeks to discover how and why anyone can contract each of these diseases, at any time. Fred is interesting to him, if at all, as one more piece of evidence for an illness in one of his categories, a piece of evidence which his theory needs to find room for. So it is with all the sciences; and so much for the yearning to get neurophysiologists to tell us what process in the brain correlates with X's thought at t that he must buy more toothpaste, or with anyone's thought at any time that he or she must buy more toothpaste.

In sum: the mental (by which I mean the ordinary language mental ontology) is at two removes from the physical (by which I mean the referents of neurophysiological descriptions of sequences of cerebral processes). To discuss the relation of everyday concepts with physical ones, we have first to give up talking about an individual occurrence of ϕ-ing at time t, and turn to talk instead about the capacity to ϕ; then we must specify in the alternative and scientific vocabulary of psychology what state or states we are asking about. For example, the original question 'How did John remember my telephone number?' will be transmuted, *via* 'How do people remember telephone numbers?' into 'How is it that people can store a string of numbers in medium-term memory?' The next stage is to learn about the psychological-physical correlation by learning how the psychological process (for it *is* now a process) can be unfolded into a structure, each element of which—or the constituents of each element of which—is embodied in some complex neural processes. Above all, the step of translating from the mental to the psychological cannot be omitted. To suppose that it could be omitted would be as gross a mistake as it would be to wonder why particle physics cannot provide lawlike correlations between the entities and processes of its domain and objects like cuckoo-clocks, cameras, or copies of *Candide*.[1]

[1] Earlier versions of this paper were read at Oxford and MacMaster, and it has been much improved by the criticisms it received at both places. I am very grateful also for the comments of Tom Nagel and Professor J. W. N. Watkins.

St Hilda's College, Oxford

REFERENCES

BORNSTEIN, B. [1963]: 'Prosopagnosia', in L. Halpern (*ed.*), *Problems in Dynamic Neurology*. Jerusalem.

BRAIN, W. R. [1941]: 'Visual Disorientation with Special Reference to Lesions of the Right Cerebral Hemisphere', *Brain*. 64, pp. 244–72.

BRUNER, J., GOODNOW, J. J. and AUSTIN, G. A. [1956]: *A Study of Thinking*. London.

DAVIDSON, D. [1970]: 'Mental Events' in L. Foster and J. W. Swanson (*eds.*), *Experience and Theory*. London.

DENNETT, D. C. [1975]: 'Why the Law of Effect Will not Go Away', *Journal for the Theory of Social Behaviour*, 5, pp. 169–87.

DENNETT, D. C. [1978]: 'Why You Can't Make a Computer that Feels Pain', *Synthèse*, 38, pp. 414–56.

DIMOND, S. J. [1972]: *The Double Brain*. London.

DUNHAM, P. J. [1977]: *Experimental Psychology: Theory and Practice*. London.

FOLTZ, E. and WHITE, L. [1962]: 'Pain "Relief" by Frontal Cingulotomy', *Journal of Neurosurgery*, 19, pp. 89–100.

GALL, F. J. and SPURZHEIM, J. [1810–1818]: *Anatomie et Physiologie du Système Nerveux en Général et du Cerveau en Particulier*. Paris.

GARDNER, H. [1977]: *The Shattered Mind.* London.

GESCHWIND, N. [1974]: *Selected Papers on Language and the Brain.* Boston Studies in the Philosophy of Science, vol. XVI, Boston.

HEATH, R. G., MONROE, R. R. and MICKLE, W. A. [1955]: 'Stimulation of the Amygdaloid Nucleus in a Schizophrenic Patient', *American Journal of Psychiatry*, **111**, pp. 862–3.

HECAEN, H. [1978]: 'Right Hemisphere Contribution to Language Functions', in P. A. Buser and A. Rougeul-Buser (*eds.*), *Cerebral Correlates of Conscious Experience.* Amsterdam.

HINSHELWOOD, J. [1900]: *Letter-, Word-, and Mind-Blindness.* London.

LASSEN, N. A., INGVAR, D. H. and SKINHØJ, E. [1978]: 'Brain Function and Blood Flow', *Scientific American*, **239**, pp. 62–71.

LURIA, A. R. [1973]: *The Working Brain.* London.

NAGEL, T. [1965]: 'Physicalism', *Philosophical Review*, **74**, pp. 339–56.

POSNER, M. I. [1969]: 'Representational Systems for Storing Information in Memory', in: G. Talland and N. Waugh (*eds.*), *The Pathology of Memory.* New York.

SAGAN, C. [1977]: *The Dragons of Eden.* New York.

SCOVILLE, W. B. [1971]: 'The Effect of Surgical Lesions of the Brain on Psyche and Behaviour in Man', in: A. Winter (*ed.*), *The Surgical Control of Behaviour.* Illinois.

TAYLOR, C. [1967]: 'Mind-Body Identity, A Side-Issue?' *Philosophical Review*, **36**, pp. 201–13.

TEUBER, H.-L. [1955]: 'Physiological Psychology', *Annual Review of Psychology*, **6**, pp. 267–96.

TEUBER, H.-L. [1972]: 'Unity and Diversity of Frontal Lobe Functions', *Acta Neurobiologiae Experimentalis Warsaw*, **32**, pp. 615–56.

TREVARTHEN, C. [1974]: 'Analysis of Cerebral Activities that Generate and Regulate Consciousness in Commissurotomy Patients', in: S. J. Dimond and J. G. Beaumont (*eds.*), *Hemisphere Function in the Human Brain.* London.

VALENSTEIN, E. S. [1973]: *Brain Control.* New York.

WILKES, K. V. [1974]: 'Anthropomorphism and Analogy in Psychology', *Philosophical Quarterly*, **24**, pp. 126–37.

ZAIDEL, E. [1978]: 'Lexical Organisation in the Right Hemisphere', in: P. A. Buser and A. Rougeul-Buser (*eds.*), *Cerebral Correlates of Conscious Experience.* Amsterdam.

Brit. J. Phil. Sci. 31 (1980), 131–144 *Printed in Great Britain*

Bayesian Conditionalisation and the Principle of Minimum Information

by P. M. WILLIAMS

The use of the principle of minimum information, or equivalently the principle of maximum entropy, has been advocated by a number of authors over recent years both in statistical physics as well as more generally in statistical inference.[1] It has perhaps not been sufficiently appreciated by philosophers, however, that this principle, when properly understood, affords a rule of inductive inference of the widest generality.[2] The purpose of this paper is to draw attention to the generality of the principle. Thus the Bayesian rule of conditionalisation, as well as its extension by R. C. Jeffrey, will be exhibited as special cases. General conditions under which it yields a unique prescription will also be studied. Detailed treatment will be restricted to the finite-dimensional case but an outline of the general case is given in the Appendix.

The underlying idea of maximum entropy inference is this. Suppose P to be a probability distribution assigning probabilities $p_1, \ldots, p_n$ to n mutually exclusive and jointly exhaustive events. Then the information-theoretic *entropy* of the distribution is defined by

$$S(P) = -\sum_{j=1}^{n} p_j \log p_j,$$

where it is to be understood that any term in the summation for which $p_j = 0$ vanishes. $S(P)$ can be thought of as measuring the 'missing information' or 'uncertainty' associated with the distribution. In particular, $S(P)$ assumes its maximum value for the uniform distribution and its minimum value for any distribution concentrated on a single event. The *principle of maximum entropy* states that the probability distribution appropriate to a given state of information is one that maximises entropy subject to the constraints imposed by the information given.

Suppose it to be a question of assigning probabilities to the outcomes of the throw of a loaded die where, specifically, the information available

---

[1] See, for example, Jaynes [1957] and Kullback [1959]. An extensive bibliography may be found in Guiaşu [1977].

[2] By a rule of inductive inference is meant a rule for passing from a prior probability distribution to a posterior distribution in the light of new information. The choice of an initial prior is a separate question. On this see, for example, Jaynes [1968].

Received 15 *December* 1978

makes it appropriate to require the expected score to be 4·5 on the throw in question. Then, if no other constraint is specified, the principle of maximum entropy prescribes the distribution

$$p_j = A\theta^j \qquad (j = 1, \ldots, 6)$$

where the normalising factor A and the parameter θ are to be determined by the conditions

$$\sum_{j=1}^{6} A\theta^j = 1 \qquad \sum_{j=1}^{6} Aj\theta^j = 4{\cdot}5$$

This may be regarded as the most 'spread out' of all distributions consistent with the constraint.

In the absence of any constraint other than that the sure event should have unit expectation, the principle of maximum entropy yields the uniform distribution for a finite sample space. It would therefore appear to be open to whatever objections can be raised against the classical principle of indifference as a principle for establishing prior distributions. According to the present viewpoint, however, this is not the correct way of viewing the matter. It is more accurate to regard the uniform distribution as emerging from the principle of maximum entropy, in the absence of any irredundant constraint, as a result of an implicit prior decision that the uniform distribution would best express the state of information in that situation. Indeed, many derivations of the entropy function assume this explicitly at the outset by requiring that any adequate measure must assume its maximum value for the uniform distribution.[1] This requirement is no longer appropriate in the presence of another prior distribution. In that case, whatever distribution has been chosen as prior should be the one to emerge from an application of the principle of maximum entropy in the absence of any further constraint. If the principle is to have this consequence, it is necessary to generalise the entropy concept so as to be applicable relative to any prior distribution. A generalisation of this type is well known in the case of continuous distributions. Its need in the finite case, however, has been obscured by the fact that a sample space can generally be chosen over which the uniform distribution is an adequate expression of prior opinion.[2]

Suppose that a prior distribution P^0 is given. Then the required generalisation is given by the expression

---

[1] Khinchin [1957], pp. 9–13.

[2] Even so, ambiguities can arise. Suppose that probabilities are to be assigned to the $n+1$ possible proportions of successes in a sequence of n binary trials, subject to a constraint on the expected proportion. Clearly the principle, as stated, will give a different result depending on whether it is applied to the space of $n+1$ proportions or the space of 2^n sequences.

$$-\sum_{j=1}^{n} p_j \log(p_j/p_j^0)$$

where p_j^0 is the prior probability of the jth event. This reduces to the previous expression, to within an additive constant, in the case of a uniform prior. But since the new expression is never positive, it is more meaningful to define instead the ***information in P relative to P^0*** as

$$I(P, P^0) = \sum_{j=1}^{n} p_j \log(p_j/p_j^0).$$

Every term for which $p_j = 0$ is understood to vanish whereas it is convenient to set $I(P, P^0) = +\infty$ if $p_j^0 = 0$ in any remaining term for which p_j is non-zero.[1]

The following property is fundamental:

$$I(P, P^0) \geqslant 0 \textit{ with equality iff } P = P^0.$$

Proof. The result does not depend on the base of the logarithms which may therefore be assumed to be natural for simplicity. If any p_j^0 vanishes whilst the corresponding p_j does not, the expression diverges and is certainly strictly positive. Suppose, therefore, that $p_j^0 = 0$ implies $p_j = 0$ for all $= 1, \ldots, n$. Since, for any positive real number x,

$$x \log x - x + 1 \geqslant 0 \textit{ with equality iff } x = 1,$$

we obtain

$$I(P, P^0) = \sum_j p_j^0 \left\{ \frac{p_j}{p_j^0} \log \frac{p_j}{p_j^0} - \frac{p_j}{p_j^0} + 1 \right\} \geqslant 0$$

where the summation need extend only over those j for which $p_j \neq 0$ and for which, consequently, $p_j^0 \neq 0$. It follows that, for these j, $I(P, P^0) = 0$ implies $p_j = p_j^0$. For the remaining j for which $p_j^0 = 0$, we have already seen that necessarily $p_j = p_j^0$, otherwise $I(P, P^0)$ would diverge. This completes the proof.

---

[1] Expressions of this type, or their extensions to the continuous case, are already studied in Good [1950], Kullback and Leibler [1951], Savage [1954] and Lindley [1956]. For more such references see Kullback [1959], ch. 1. Mention of earlier work by A. M. Turing is made in Good [1950]. A derivation of the expression $I(P, P^0)$ as the unique measure of relative information satisfying suitably modified forms of the Shannon–Khinchin conditions may be found in Hobson [1971], Appendix A. It is possible, however, to have reservations concerning such derivations on the grounds that they require that entropy, or information, should be, in a certain sense, additive. From the point of view of maximum entropy inference, however, this goes beyond the intuitions which the adequacy conditions are intended to express. For these concern themselves only with the ordering of distributions—with the idea, for example, of one distribution deviating more or less than another from the prior distribution. Additivity is, no doubt, a convenient property but it is not required for maximum entropy inference. Any strictly monotonic function of the usual expression would serve equally well. It would seem preferable, therefore, if an argument for the unique entropic ordering of distributions could be found which restricted itself to adequacy conditions formulated in terms of order alone.

It follows from this that, in the absence of any constraint other than that of correct normalisation, the information in P relative to P^0 is uniquely minimised by the prior distribution as was to be required. Any other distribution contains positive information. In the case of a constraint specifying the expected score on the throw of a die, for example, the new minimum information distribution is given by

$$p_j = Ap_j^0\theta^j \qquad (j = 1, \ldots, 6)$$

where A and θ are to be determined as before, though now taking into account the prior probabilities. As in the Bayesian case, the posterior probability is found by multiplying the prior probability by a numerical factor.

The generalised principle of maximum entropy or, better, the ***principle of minimum information*** can now be formulated as follows:

Principle of Minimum Information: *Given the prior distribution P^0, the probability distribution P appropriate to a new state of information is one that minimises $I(P, P^0)$ subject to whatever constraints the new information imposes.*[1]

It is important to emphasise that the principle of minimum information, in this form, is not a principle for setting up prior distributions. It is, rather, a general principle of probability dynamics. It seeks to answer the question how to ***modify*** a probability distribution, in the light of new information, in the most conservative way. It is inapplicable in the absence of a prior distribution.[2]

As a special case, suppose that a prior distribution P^0 is given and that new evidence establishes some proposition E in the domain of P^0 with certainty. Then the new distribution incorporating this information must attribute unit probability to this event. It follows from the principle of minimum information that, provided $P^0(E)$ is non-zero, the new distribution is given, for any F, by

---

[1] A formulation of this principle, taking explicit account of the prior distribution, already occurs in Good [1963] as the 'Principle of Minimal Discriminability'. Good states that, according to his interpretation, its purpose is to generate null hypotheses concerning physical probability distributions which are to be tested by experiment. This contrasts, in form of expression at least, with the present interpretation according to which the purpose of the principle is to assist in the rational modification of beliefs.

[2] According to the present interpretation, the probabilities emerging from the principle of minimum information are not conditional probabilities associated with the prior distribution but unconditional probabilities of a new and entirely different distribution, unrelated to the prior distribution by the normal 'synchronic' probability calculus. This is to be understood even in the case corresponding to Bayesian conditionalisation. If this is accepted, objections of the type raised by Friedman and Shimony [1971] are not applicable.

$$P(F) = \frac{P^0(EF)}{P^0(E)} = P^0_E(F).$$

The most meaningful way of establishing this is to verify first that, for any distribution for which $P(E) = 1$, the information in P relative to P^0 is given by

$$I(P, P^0) = I(P_E, P^0_E) - \log P^0(E)$$

This is minimised when $P_E = P^0_E$ with the increase in information given by the remaining constant term $-\log P^0(E)$. The posterior distribution is thus uniquely determined given that $P(E) = 1$. Thus the Bayesian rule of conditionalisation is a special case of the principle of minimum information.[1]

It has been suggested by R. C. Jeffrey [1965], however, that the rule of conditionalisation is, strictly speaking, limited in its applicability inasmuch as it requires new information to establish some proposition with certainty. In practice there is always room for doubt. Suppose therefore that new evidence establishes some proposition E in the domain of the prior distribution P^0 only with probability q, where q may be supposed close to unity, though this is not essential. Jeffrey has proposed that the posterior distribution be given by

$$P(F) = qP^0(F|E) + \tilde{q}P^0(F|\bar{E}) \qquad (\tilde{q} = 1-q).$$

This is the solution to a special case of the general question raised by Jeffrey as follows[2]: 'Given that a passage of experience has led the agent to change his degrees of belief in certain propositions $E_1, E_2, \ldots, E_m$ from their original values

$$P^0(E_1), P^0(E_2), \ldots, P^0(E_m)$$

to new values,

$$P(E_1), P(E_2), \ldots, P(E_m)$$

how should these changes be propagated over the rest of the structure of his beliefs?'

This is the sort of question which the principle of minimum information is designed to answer. In the special case just considered, where one takes

[1] It must be admitted, however, that this has only been demonstrated in the case where $P^0(E)$ is strictly positive whereas the Bayesian rule, to replace $P^0(F)$ by $P^0(F|E)$ when E has become certain, is asserted even when $P^0(E) = 0$, assuming the conditional probabilities to be defined independently. When $P^0(E)$ vanishes, the information in any distribution assigning positive probability to E necessarily diverges and no choice can be made on this basis. This is as it should be. Relative information has been defined only for unconditional distributions, which say nothing about the relative probabilities of events of probability zero. To deal adequately with this case it would be necessary to define the relative information of conditional distributions or, at least, the corresponding ordering.

[2] Jeffrey [1965], p. 157 adapted to the present notation

account of only a single event or proposition at a time, the principle of minimum information yields the solution advocated by Jeffrey. The same is true in the case of several mutually exclusive events. To prove this the case of a single event, it is routine to verify that provided $0 < q < 1$ we have, for any distribution P for which $P(E) = q$,

$$I(P, P^0) = qI(P_E, P^0_E) + \bar{q}I(P^0_{\bar{E}}, P^0_{\bar{E}}) + \text{constant}$$

where again

$$P_E(F) = \frac{P(EF)}{P(E)} \quad \textit{etc.}$$

Clearly this is minimised by separately minimising the functions on the right. Thus we obtain Jeffrey's result $P = qP^0_E + \tilde{q}P^0_{\bar{E}}$. It is straightforward to extend this to any number of events provided they are mutually exclusive. This is in fact the most general case treated by Jeffrey. In the case where the events are not necessarily mutually exclusive, the principle of minimum information gives new results. Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be the space of possibilities. Suppose the events $E_1, \ldots, E_m \subseteq \Omega$ are constrained to have new probabilities $q_1, \ldots, q_m$ and, for convenience, write E_0 for the sure event with probability $q_0 = 1$. Assume, for simplicity, that P^0 is strictly positive and that there exists a strictly positive distribution satisfying the constraints. This means, in particular, that none of the new probabilities is 0 or 1. Then, defining the array $\{\chi_{ij}\}$ by

$$\chi_{ij} = \begin{cases} 1 & \text{if } \omega_j \in E_i \qquad (i = 0, \ldots, m) \\ 0 & \text{otherwise} \qquad (j = 1, \ldots, n) \end{cases}$$

the solution is given by

$$p_j = p_j^0 \exp\left\{\sum_{i=0}^{m} \lambda_i \chi_{ij}\right\} \qquad (j = 1, \ldots, n)$$

with the parameters $\lambda_0, \ldots, \lambda_m$ determined by the constraints

$$q_i = \sum_{j=1}^{n} \chi_{ij} p_j \qquad (i = 0, \ldots, m).$$

There is an interesting way in which the transformation of probability distributions by the principle of minimum information appears to differ, in general, from the special case of Bayesian conditionalisation. For, in the latter case, it makes no difference whether one conditionalises successively on E_1 and E_2, in either order, or directly on the joint event E_1E_2. On the face of it, this is not true of the general case.[1] If two constraints are applied successively, there is no reason, in general, why the first should remain satisfied when information is minimised subject to the second. The reason why it is unnecessary to maintain the first constraint explicitly in the Bayesian case is that no subsequent application of the principle of minimum information can ever reverse the decision to assign probability 1 to a given event (unless only one distribution satisfies the later constraint when no further principle is needed anyway). If then one regards the Bayesian case as one in which earlier constraints are implicitly maintained, the position is not so different in the general case. For suppose that, in the situation just treated, we define a sequence of probability distributions $\{P^i\}$ by requiring that for each $i = 1, \ldots, m$

---

[1] *Cf.* Jeffrey [1965], pp. 162–3.

$$P^i \textit{ minimises } I(P, P^{i-1}) \textit{ subject to } P(E_h) = q_h \textit{ for all } h = 1, \ldots, i.$$

At each stage all earlier constraints are explicitly maintained. Then it is not difficult to show that we arrive finally at the same distribution as we found before by applying all the constraints at once, which also means that the order is unimportant. This is not trivial since finally we have reached the same probability distribution by applying the same constraint to two different priors, namely P^{m-1} and P^0. In this respect the principle of minimum information behaves as well as one could expect. We shall return to this question, in greater generality, later.

It is time now to deal with the question of the effectiveness of the principle, that is with the existence and uniqueness of information-minimising distributions. The general situation is this. A prior distribution is assumed but, for one reason or another, it may no longer be an adequate expression of opinions. The new distribution should belong to a certain subset $\mathscr{C}$ of the set $\mathscr{P}$ of possible distributions. In the finite-dimensional case $\mathscr{C}$, like $\mathscr{P}$, is a subset of $\mathbf{R}^n$. $\mathscr{C}$ will normally contain many members and a choice must be made between them. The principle of minimum information prescribes the distribution in $\mathscr{C}$ that minimises $I(P, P^0)$. In principle, $\mathscr{C}$ could be an arbitrary subset of $\mathscr{P}$ and there is no guarantee that such a distribution exists. Let us see what can go wrong.

(i) $\mathscr{C}$ is empty: the constraints are inconsistent.

(ii) $\mathscr{C}$ is non-empty but $I(P,P^0)$ is infinite for all P in $\mathscr{C}$. This will occur, in the finite-dimensional case, if and only if $\mathscr{C}$ requires that a positive probability be assigned to an event of prior probability zero.

(iii) Now $\mathscr{C}$ contains distributions with finite information relative to P^0 but no minimal element. For every distribution in $\mathscr{C}$ there is another with strictly smaller information. This will arise if it is required that $P(E) > q$, where $0 < P^0(E) \leqslant q < 1$, or that $I(P, P^0) > \alpha$ for suitable $\alpha \geqslant 0$. (If $\alpha = 0$, this only says that P should be different from P^0.)

(iv) $\mathscr{C}$ contains minimal elements, with finite information relative to P^0, but more than one. *Example*: except in certain trivial cases, the constraint that $I(P,P^0) = \alpha$ for some finite $\alpha > 0$.

We shall examine these possibilities in turn beginning with the last.

Case (iv) will never arise if $\mathscr{C}$ is *convex*. For, with P^0 fixed, $I(P, P^0)$ is a *strictly* convex function of P. Thus if P^1 and P^2 are two distinct distributions with the same finite information relative to P^0, any proper convex combination

$$\lambda P^1 + (1-\lambda)P^2 \qquad (0 < \lambda < 1)$$

has strictly smaller information. The case of convex constraints appears to be the rule rather than the exception. Certainly constraints on expectation values are of this type, however many random quantities they concern and whether they are expressed by linear equalities or weak or strong inequalities.

Case (iii) will never arise if $\mathscr{C}$ is a *closed* subset of $\mathscr{P}$ (usual topology). So far as constraints on expectation values are concerned, this means that, in general, strong (strict) inequalities are to be avoided. The fact that the principle fails to make a definite recommendation in such cases is surely a merit not a defect. Consider the simplest case where $P^0 = (\frac{1}{2}, \frac{1}{2})$, the uniform distribution over two possibilities, heads and tails say. If the constraint demands only that $P(H) > \frac{1}{2}$, it is hard to see how any principle could reasonably be expected to yield a prescription.

As was already observed, case (ii) arises if subsequent evidence requires that a strictly positive probability be assigned to an event of prior probability zero. Then it will be impossible to distinguish between distributions on the basis of their relative information as it has been defined. To obtain definite results here would require a more delicate method of comparison. An extension of the definition in this direction would certainly be of interest, though beyond the scope of this paper.[1] On the other hand it is relevant to observe that the principle itself has no inherent tendency to lead to more concentrated distributions. On the contrary, the principle is intrinsically conservative. In the case of a convex constraint, for instance, the supports of the prior and posterior distributions will coincide unless the constraint explicitly requires it to be otherwise, that is unless there is no distribution with the same support as the prior that satisfies the constraint. To this extent the principle itself contributes nothing towards the difficulties arising under (ii).

Lastly, there is nothing to be done in case (i) except to re-examine the constraints to see where one went wrong. This is the province of the 'synchronic' probability calculus.

---

[1] *Cf.* p. 135, n. 1.

In summary, the principle of minimum information yields a unique prescription for all closed convex constraints satisfied by at least one distribution having finite information relative to the given prior.

It is worth returning briefly to the question of the significance of the order in which constraints are applied. It was mentioned before that, in the case of Bayesian conditionalisation, it makes no difference whether one conditionalises successively on two events E_1 and E_2, in either order, or directly on the joint event E_1E_2. It was observed that a corresponding result holds in the more general case contemplated by Jeffrey provided that, at each stage, all earlier constraints are explicitly maintained. This amounts to dealing with a sequence of successively stronger constraints. The result claimed in that case was that the outcome of applying the constraints one by one is the same as applying them all at once, namely just the last. In fact it seems reasonable to expect the order to be unimportant only in such a situation. Suppose, therefore, that we have a sequence of successively stronger constraints where, indeed, without loss of generality we may suppose there to be just two: $\mathscr{C}_1 \supseteq \mathscr{C}_2$. Let P^0 be given and suppose that P^1 alone, amongst distributions in $\mathscr{C}_1$, has minimal information relative to P^0. Suppose again that P^2 alone, amongst distributions in $\mathscr{C}_2$, has minimal information relative to P^1. The question at issue is whether P^2 is the same distribution as would have been reached by applying the constraint $\mathscr{C}_2$ directly to P^0. In fact a sufficient condition for this to happen is that $\mathscr{C}_1$ should be an *affine* constraint. Then, provided $\mathscr{C}_2$ is also affine, the result can be extended to a third constraint, and so on.

An affine constraint is one that contains every probability distribution on a line through any two of its members. That is to say, along with any P_0 and P_1, it contains

$$P_\lambda = (1-\lambda)P_0 + \lambda P_1$$

for every real number λ for which P_λ remains non-negative. Constraints on expectation values are of this type. Thus if $\bar{x}_1, \ldots, \bar{x}_m$ are any real numbers and $X_1, \ldots, X_m$ are any real-valued functions on $\Omega = \{\omega_1, \ldots, \omega_n\}$, the set of probability distributions $P = (p_1, \ldots, p_n)$ such that for each $i = 1, \ldots, m$

$$\sum_{j=1}^{n} p_j X_i(\omega_j) = \bar{x}_i$$

forms a (possibly empty) affine constraint. Indeed, as is well known, this is the most general example in finite dimensions.[1]

Suppose then that $\mathscr{C}$ is an affine constraint containing at least one distribution having finite information relative to P^0. Then, necessarily, there is a unique distribution, P^1 say, in $\mathscr{C}$ having minimal information relative to P^0. Furthermore, for any distribution P in $\mathscr{C}$,

---

[1] Previously we were dealing with the special case of an affine constraint determined by $\{0,1\}$—valued random quantities.

$$I(P, P^0) = I(P, P^1) + I(P^1, P^0).$$

Thus, the amounts of information in any distribution belonging to $\mathscr{C}$, relative to P^0 and P^1, differ by the same constant amount, provided they are finite. Otherwise they diverge simultaneously. This means that the functions $I(-, P^0)$ and $I(-, P^1)$ induce the same orderings in $\mathscr{C}$ and, *a fortiori*, in any stronger constraint that might be applied subsequently.

It is not possible to extend this result beyond the class of affine constraints in a straightforward way. Certainly it fails for convex constraints in general. Whether or not this should be considered a defect of the principle will depend on what alternatives exist. Nevertheless, it seems reasonable to demand of any acceptable alternative that it should possess the property in question for the restricted, but important, class of affine constraints at least. It is not obvious, however, that any principle, other than that of minimum information, meets even this limited demand. Except in the trivial[1] case $n \leqslant 3$, it is certainly not met by minimising (the square of) the Euclidean distance from the prior:

$$\sum_{j=1}^{n} (p_j - p_j^0)^2.$$

On the other hand, that principle might already be excluded on the grounds that it fails to yield the Bayesian rule of conditionalisation and has no natural infinite-dimensional generalisation. An improvement, in the latter respect at least, is afforded by the principle prescribing a distribution that minimises the distance from the prior in the sense of the metric[2]:

$$D(P, P^0) = \sup \{|P(E) - P^0(E)| : E \subseteq \Omega\}.$$

The Bayesian rule and Jeffrey's generalisation are at least optimal from this point of view, though by no means uniquely so even in elementary cases. In respect of its treatment of a sequence of successively stronger affine constraints, on the other hand, this principle fares no better in general (in cases where it succeeds in prescribing a unique posterior) than the principle based on the Euclidean metric. It is perhaps premature to conjecture that the principle of minimum information is unique in possessing the property in question. It is nonetheless possible that, by studying such transformational characteristics of various principles, a more satisfactory explanation of the peculiar reasonableness of the principle of minimum information will be found.

---

[1] If $n < 3$, any principle whatever has the property in question for a decreasing sequence of affine constraints, provided only that it preserves the prior whenever the prior already satisfies the constraint.

[2] This is essentially the principle discussed in the paper of S. May and W. Harper [1976] which was kindly brought to my attention by one of the referees. May and Harper mention the principle of minimum information as a possible 'minimum change' principle, but choose instead the principle based on a version of the supremum metric for detailed study.

APPENDIX

Suppose that the events to which probabilities are to be assigned can be represented by a field $\mathscr{F}$ of subsets of a non-empty set Ω of arbitrary cardinality. A probability distribution over $\mathscr{F}$ is a non-negative, finitely additive[1] real-valued function on $\mathscr{F}$ such that $P(\Omega) = 1$. Let $\mathscr{P}$ be the set of all probability distributions over $\mathscr{F}$ and let P and P^0 be two such distributions. For any finite partition $\mathscr{E} = \{E_1, \ldots, E_n\}$ of Ω into sets belonging to $\mathscr{F}$, the information in P relative to P^0 with respect to $\mathscr{E}$ can be defined by

$$I_{\mathscr{E}}(P,P^0) = \sum_{j=1}^{n} P(E_j) \log \frac{P(E_j)}{P^0(E_j)}$$

which is to be understood in the same way as before when any of the terms in the summands vanish.

When confronted with a variety of finite schemes, as in the infinite case, it is only reasonable, if the aim is to minimise information, to assume no less than the worst case, though nothing positively worse. Thus it is natural to define the information in P relative to P^0 as the supremum over finite partitions:

$$I(P, P^0) = \sup \{I_{\mathscr{E}}(P, P^0)\}.$$

It is worth remarking that $I_{\mathscr{E}}(P, P^0)$ can only increase as the partition is further refined.

In this way $I(P, P^0)$ is defined for every P in $\mathscr{P}$ provided we admit the value $+\infty$. Clearly $-\infty$ is not possible value. In fact it is clear that

$$I(P, P^0) \geqslant 0 \textit{ with equality if and only if } P = P^0.$$

The distributions P for which $I(P, P^0)$ assumes a finite value will be said to belong to the *effective domain* of the function $I(-, P^0)$. A necessary, though generally not sufficient, condition for $I(P, P^0)$ to be finite is that P should be *absolutely continuous* with respect to P^0. That is to say, for any $\epsilon > 0$ there should exist a $\delta > 0$ such that, for every E in $\mathscr{F}$,

---

[1] The present approach endorses the viewpoint of de Finetti (see [1974], ch. 6 for example) that it is better not to assume at the outset that all probability distributions are countably additive, but to assume only those properties that follow from the meaning of probability itself—which do not include countable additivity—and to introduce the latter as a special assumption only when particular circumstances justify.

$$P^0(E) < \delta \textit{ implies } P(E) < \epsilon.$$

In this case, if P and P^0 are countably additive over a σ-field, a Radon–Nikodym derivative dP/dP^0 exists and the equivalent integral expression

$$I(P, P^0) = \int_\Omega \log(dP/dP^0)\, dP$$

can be obtained (Guiaşu [1977], ch. 2).

In general, without making the assumption of countable additivity, it can be shown that

$$I(-, P^0) \textit{ is strictly convex over its effective domain}$$

for any prior distribution P^0. Furthermore, if we endow $\mathscr{P}$ with the relativised product topology by considering $\mathscr{P}$ as a subset of the product $\mathbf{R}^{\mathscr{F}}$, it follows that

$$I(-, P^0) \textit{ is lower semi-continuous}$$

for any prior distribution P^0.

From the last result, together with the well-known compactness of $\mathscr{P}$, it follows that $I(-, P^0)$ assumes its minimum value on any closed subset $\mathscr{C}$ of $\mathscr{P}$. That is, there exists a distribution P^1 in $\mathscr{C}$ such that

$$I(P^1, P^0) = \inf\{I(P, P^0) \colon P \in \mathscr{C}\}.$$

Furthermore, if $\mathscr{C}$ is convex and includes a distribution having finite information relative to P^0, strict convexity of $I(-, P^0)$ implies that P^1 is unique. Thus, in general, the principle of minimum information yields a unique prescription for all closed convex constraints satisfied by at least one distribution having finite information relative to the given prior.

It was observed in the finite-dimensional case that the principle has no inherent tendency to lead to either more or less concentrated distributions. In general this is expressed by the mutual absolute continuity of prior and posterior distributions. Let us assume that P^1 has minimal finite information over $\mathscr{C}$ relative to P^0. Then certainly P^1 is absolutely continuous with respect to P^0 since $I(P^1, P^0)$ is finite. Conversely, if $\mathscr{C}$ is convex, P^0 is absolutely continuous with respect to P^1 unless there is no distribution in $\mathscr{C}$ having finite information relative to P^0 and with respect to which P^0 is absolutely continuous. Recalling that no purely finitely additive distribution is absolutely continuous with respect to a countably additive distribution, it follows that from a countably additive prior the principle can only lead to a countably additive posterior.[1] On the other hand, the principle leads from a purely finitely additive prior to a countably additive posterior only if there is no purely finitely additive distribution in $\mathscr{C}$ (assumed to be convex) having finite information relative to P^0. So that although countably additive distributions are 'absorbing' in this sense, the principle displays no inherent attraction towards them.

---

[1] By a purely finitely additive distribution is meant one that is finitely but not countably additive.

It is of interest to note that the extension of distributions by means of countable additivity has no effect on their relative information. If P, P^0 are countably additive distributions over the field $\mathscr{F}$ and $\hat{P}$, $\hat{P}^0$ are their unique countably additive extensions to the σ-field generated by $\mathscr{F}$, then

$$I(\hat{P}, \hat{P}^0) = I(P, P^0)$$

where it is to be understood that the two amounts of information are calculated by means of finite partitions of the extended and restricted fields, respectively.

All that was said before in the finite case concerning Bayesian conditionalisation as an example of the principle of minimum information holds good in general. The same is true of Jeffrey's rule when there are finitely many events whose posterior probabilities are prescribed. In the general case, however, Jeffrey's problem may be formulated with respect to the prescription of infinitely many posterior probabilities. Provided some distribution having finite information relative to the given prior satisfies this prescription, the principle of minimum information again leads to a unique solution.

The question of the significance of the order in which constraints are applied requires more delicate and extended discussion in the general case. This will be dealt with elsewhere.

The University of Sussex

REFERENCES

de Finetti, B. [1974]: *Theory of Probability*. London: John Wiley & Sons.

Friedman, K. and Shimony, A. [1971]: 'Jaynes's Maximum Entropy Prescription and Probability Theory', *J. Statist. Phys.* **4**, pp. 381–4.

Good, I. J. [1950]: *Probability and the Weighing of Evidence*. London: Charles Griffin & Co. Ltd.

Good, I. J. [1963]: 'Maximum Entropy for Hypothesis Formulation, especially for Multi-dimensional Contingency Tables', *Ann. Mat. Statist.* **34**, pp. 911–34.

Guiaşu, S. [1977]: *Information Theory with Applications*. New York: McGraw-Hill, Inc.

Hobson, A. [1971]: *Concepts in Statistical Mechanics*. New York: Gordon and Breach, Inc.

Jaynes, E. T. [1957]: 'Information Theory and Statistical Mechanics', *Phys. Rev.* **106**, pp. 620–30; **108**, pp. 171–90.

Jaynes, E. T. [1968]: 'Prior Probabilities', *IEEE Trans. Syst. Science and Cybernetics*, SSC–4, pp. 227–41.

Jeffrey, R. C. [1965]: *The Logic of Decision*. New York: McGraw-Hill, Inc.

Khinchin, A. I. [1957]: *Mathematical Foundations of Information Theory*. New York: Dover Publications, Inc.

Kullback, S. and Leibler, R. A. [1951]: 'On Information and Sufficiency', *Ann. Math. Statist.* **22**, pp. 79–86.

Brit. J. Phil. Sci. 31 (1980), 145–163 *Printed in Great Britain*

Models in Physics*

by MICHAEL REDHEAD

1 *Introduction*
2 *The Genesis of Models*
3 *Models and Analogies*
4 *Models and Approximations*
5 *The Uses of Models*
  (*i*) *The Probing of Theories*
  (*ii*) *The Heuristic Rôle of Models*
  (*iii*) *The Testing of Theories*
6 *Floating Models*
7 *Conclusion*

1 INTRODUCTION

The object of the present paper is to consider the rôle of models in modern theoretical physics. We want to discuss the circumstances in which models arise and the uses to which they are put, comprising in brief summary the probing of theories, the discovery of new theories and the empirical testing of theories. In the first place we want to describe the way the term model is actually used by physicists and in particular how it is related to the idea of approximation. We shall then want to expose as clearly as possible the logical structure of the arguments used by physicists when they employ models for various different purposes. We shall be at pains to point out the logical gaps in these arguments and thereby to provide a critique of some vital aspects of modern physics. Some of the very extensive literature on models in science[1] has been concerned with attempting to justify the heuristic rôle of models in producing predictive theories in the sense of theories which are devised to explain or comprehend one group of phenomena and succeed in giving successful predictions about some other quite different phenomena. This involves typically a realist approach to theories combined with a technical 'logic' of analogy.[2] The sense of analogy employed in these arguments is described in section 3 *below.*

---

* A previous version of this paper was read at the Annual Conference of the British Society for the Philosophy of Science held at the University of Sussex in September 1977. I am grateful for helpful discussions and comments provided by Mary Hesse, Noretta Koertge, Heinz Post, John Dorling, Donald Gillies and Allan Franklin. Errors and confusions that remain are my own responsibility.

[1] For a survey of the literature with an extensive bibliography Leatherdale [1974] may be consulted.

[2] See in particular Hesse [1963] and [1974].

However, it does not appear to play a central rôle in modern theoretical physics. Furthermore the rôle of analogy in characterising satisfactory explanations as ones which relate the unknown or less familiar to the known or familiar[1] is no longer conspicuous. Now, while admitting the importance and interest of such general questions as 'How is predictive science possible?' or 'What characterises a satisfactory explanation in science?', we want to claim as a legitimate activity for philosophy of science the humbler activity of *describing* what is actually done by scientists, with a view of course to discussing whether what is done is in fact justifiable on any acceptable canons of deductive or inductive logic. If the practice of science involves an irreducibly intuitive element we want to pin down and locate such elements as accurately as possible in the general structure of scientific activity.

We shall be concerned with theoretical models[2] in the sense used by Achinstein in his [1968]. Achinstein lists five distinguishing characteristics of a theoretical model:

(1) It is a set of assumptions about some object or system.

(2) These assumptions attribute an inner structure, composition or mechanism, which manifests itself in other properties exhibited by the object or system.

(3) These assumptions are treated as a simplified approximation useful for certain purposes.

(4) The model is proposed in the framework of some more basic theory or theories.

(5) The model may display an analogy between the object or system described and some other object or system.

In the context of the realist view of theories espoused by Achinstein the first criterion is also a necessary condition for characterising a theory, while the second condition, while not necessary for a theory, is certainly true of many theories. It is with the third condition that the distinction between theories and theoretical models is revealed. In justificationist terms we may say that we believe our theories to be true, even at a tentative level. For Ackinstein a theoretical model is definitely believed to be false.[3]

---

[1] See Campbell [1957].

[2] The terminology appears to be due originally to Black—see his [1962], p. 226. Hesse in her [1963] uses the locution model$_1$ for this concept.

[3] For a Popperian we cannot distinguish theories from models in terms of degrees of belief. The distinction can now be drawn in the following way. Theories for Popper arise in response to problem situations. A model *M* relative to a theory *T* fails to provide an adequate solution for the total problem situation including background knowledge. Models may provide adequate accounts of some aspects of both the original problem situation and of the novel predictions which allow the theory to be severely tested.

2 THE GENESIS OF MODELS

But let us turn to the fourth of Achinstein's conditions. This may be taken in two senses: (*a*) Suppose a theory T has been proposed satisfying our first two conditions for a theoretical model. *T* may be too complicated to 'solve' exactly, *i.e.* in drawing out the empirical predictions for comparison with experiment. In such a case a simplified model *M* may be employed which may allow of exact solution, or at any rate present a more tractable structure than *T*. *M* plays the rôle of an ***impoverished*** theory, the important ingredient being that *M* and *T* logically contradict one another, so that we believe *M* to be false insofar as we believe *T* to be true. (*b*) A theory may be incompletely specified in the sense that certain general constraints of symmetry and so forth may be imposed but this may still allow considerable latitude in selecting the detailed structure of the theory. For example in axiomatic field theory certain general constraints are imposed by the axioms, but there may, hopefully, be many possible fields which satisfy the axioms. The question of specific field equations and interaction terms may be left open. A model may then be introduced in the sense of filling in the missing detail. In this case the model may be said to ***enrich*** rather than impoverish a theory. We do not believe that what the model tells us is false in the sense that it contradicts what has been proposed as a complete theory—in the situation we are envisaging no such theory has been proposed. Nevertheless the guiding principle in introducing a model is once again its tractability. The possibility of the model being an exact theory is not excluded, but once this possibility is seriously entertained the model changes its status and acquires the honorific title of theory. Conversely what starts as a theory may later be seen to be inadequate in the light of empirical investigation and so be down-graded to the status of model. Relative to a more accurate theory which replaces it we now have the situation already discussed under (*a*). An example of the first change in status might be the quark model of the hadrons which has gradually acquired the current status of the quark theory. On the other hand the primitive kinetic theory of the early Maxwell would now be disparagingly referred to as the billiard ball model of a gas.

If a model is introduced in the second way we have described some properties of the model may depend only on the general constraints (to take again the example of field theory, the *CPT* theorem and the spin-statistics theorem) and are said to be model-independent. Such theorems are obviously more highly regarded by theoretical physicists than those which depend on the more detailed assumptions of a particular model. We shall return to this point when discussing the heuristic rôle of models.

3 MODELS AND ANALOGIES

We want now to comment briefly on Achinstein's fifth condition for a theoretical model—that it *may* display an analogy. Consider the relation between a model M and a theory T of which it is an impoverishment as discussed above. Let us think of T as attributing some properties $P_1\, P_2 \dots$ to a set of 'objects' which comprise the ontology of the theory. Let us denote by X those objects which possess the properties $P_1\, P_2 \dots$ Following Achinstein we distinguish two sorts of properties among the P's. Thus we write $\{P\} = \{R\} \cup \{N\}$ where $R_1\, R_2 \dots$ are the essential or (semantically) relevant properties for classifying the objects as X's and $N_1\, N_2 \dots$ are nonrelevant (or accidental or accompanying) properties. If we want to say, as Achinstein does, that M is a set of assumptions about the same objects as those referred to by T, then M must assign all the R's to these objects, and can only differ from T by changing some of the N's. Clearly in such a case T and M display an analogy in the sense that they attribute some properties which are the same, all the R's and perhaps some of the N's, and some which are different, *viz.* some of the N's. We have here a case of what Hesse in her [1963] calls a material analogy. But Achinstein wants to use the term analogy in a restricted sense, *viz.* as a relation between two essentially unlike objects or systems X and Y. To see what this means consider again our theory T. A theory T_A constitutes an analogue model for T if it changes some of the R's in such a way that the new set of R's, call it $\{R')$, is relevant for classifying the objects of the theory as Y's. An analogue model can still display material analogies in that some of the original P's are attributed by both T and T_A.[1] But the analogy may be only a formal one. In this case two theories T and T_A need not ascribe any of the same properties to the objects subsumed under the theories. All we require is a structural isomorphism so that related objects possess related properties.[1] This is the source of the very important process of cross-fertilization which plays such a conspicuous rôle in the overall growth of theoretical physics.[2] A *mathematical* model arises when we establish a formal analogy between a physical theory and some mathematical structure. This latter may often be embedded in some wider structure leading to the notion of 'surplus' structure in the mathematical formulation of a physical theory. The calculus associated with this wider structure is partially interpreted by means of the original theory T.[3] This is typically how partially interpreted theories actually arise in physics. Hesse in her [1963] explains how theories can develop by providing physical interpretation for this surplus structure and argues persuasively that the justification for obtaining a successful strongly predictive[4] theory by this manoeuvre depends on pre-theoretic material analogies with an analogue model which already provides an interpretation for the surplus structure.[5]

---

[1] We are adopting here a naive 'Platonist' view of properties as unchanging entities that can be applied in different contexts. Hesse in her [1965] recognises how meanings are changed by the context in the more sophisticated interaction view of the nature of metaphor. For purposes of exposition we shall follow Hesse's [1963] in using a rather naive formulation. Let us see if we can use our analysis to distinguish Hesse's two senses model$_1$ and model$_2$ (corresponding to Achinstein's theoretical and analogue models). Consider the example of the billiard ball model$_2$ of a gas. What distinguishes it from the model$_1$ of 'ghostly little objects'? Hesse remarks that the negative analogy which distinguishes model$_2$ as referring to billiard balls and model$_1$ as being about gases are such properties as the colour and hardness of the billiard balls. I think this choice of negative analogy is unfortunate. For example the 'colour' of a molecule might well be invoked to explain the interaction between a gas and a light wave—it belongs to the open class of neutral analogies. The negative analogy between billiard balls and molecules consists in the theoretical ascription of an inner structure to billiard balls as being composed of molecules. A cloud of minute billiard balls is not a gas of molecules just insofar as the scaled-down billiard balls are still themselves regarded as composed of molecules. Achinstein has I think correctly stressed the fact that inner structure may be semantically relevant (see his [1968], p. 24). It seems to me that the distinction between theoretical and analogue model, between model$_1$ and model$_2$, can best be brought out in terms of Achinstein's own analysis of semantical relevance. It is perhaps surprising that he omits to do this himself.

---

[1] We distinguish at this point the sense in which the term model is used in mathematical logic. A model M of a first-order theory T for example is a relational structure for which all the sentences of T express true properties about the structure M, when the latter is used to provide an interpretation of T. If we have two models M_1 and M_2 of T these are elementarily equivalent if T is complete, *i.e.*, every proposition that can be expressed in the language of T is either true or false of both M_1 and M_2. But this does not imply that M_1 and M_2 are analogous in all respects, *i.e.* M_1 and M_2 need not be isomorphic, if T is non-categorical. In this situation M_1 and M_2 stand in the relation of model to T and M_1 and M_2 stand in a relation of positive analogy to one another (elementary equivalence).

[2] Compare as just one example the use of diagram methods imported from quantum field theory into many-body problems in nuclear physics and solid state physics.

[3] These ideas are expounded in Redhead [1975].

[4] By strongly predictive Hesse means that the new theory predicts correlative laws involving new observable properties over and above those referred to by the previous theory.

[5] We comment briefly on the heuristic rôle of analogy in arriving at theories which are in Hesse's terminology weakly predictive—*i.e.* which predict new correlations between the same observable properties that the previous theory referred to. We could imagine the situation in which the old theory fails to explain some empirical correlation and we look for ways of developing the theory so as to account for the anomaly and hopefully provide sucessful novel predictions as well. Here again we can distinguish a rôle for a model$_2$ (we follow Hesse's terminology) as providing a positive heuristic in the Lakatosian sense. In our previous notation we could say that $\{R'\}$, the semantically relevant properties for classifying the objects referred to in the model as Y's, provide the hard core of a research programme based on the model. The nonrelevant properties provide the protective belt, they can be changed in accordance with the positive heuristic. Since we are interested in developing a theory of X's there will be some members of $\{R'\}$ which will be negative analogies between model$_1$ and model$_2$—they serve to distinguish model$_1$ as being a story about X's and model$_2$ a theory about Y's. The remaining members of $\{R'\}$ are positive analogies which are to be retained through all variations in the model$_1$. The

However, Achinstein's fifth condition is certainly not a necessary condition for a theoretical model and in the remainder of this paper we shall be concerned with those features of theoretical models which do not depend on analogies they may display in the sense that Achinstein uses the term analogy.

4 MODELS AND APPROXIMATIONS

First of all we want to clarify the relation between impoverishment models (from now on we shall use the term model as an abbreviation for theoretical model) and approximations. Suppose we have a theory T. Following Redhead ([1975]) we may represent T as a unary relation on some appropriate function space, each point in the function space corresponding to a logically possible coordination of the variable quantities referred to by the theory while the unary relation singles out those coordinations which T permits. In more familiar terms theories in physics are expressed *via* some system of equations whose particular solutions refer to some physically possible coordination of variables. We can consider approximations to a theory T from two points of view:

(*a*) We can look for approximate solutions of the exact equations employed in the theory. *E.g.* for the equation $dy/dx - \lambda y = 0$ we might expand our solution as a perturbation series in λ, the *nth* order approximation being just

$$y_n = 1 + \lambda x + \lambda^2 x^2/2! + \dots + \lambda^{n-1} x^{n-1}/(n-1)!$$

if we consider the boundary condition $y = 1$ at $x = 0$.

(*b*) We can look for exact solutions to approximate equations, *i.e.* we can simplify our equation before we start to solve it.

Of course the two approaches are equivalent in the sense that if we consider an approximate solution y_n for an exact theory T we can always

---

nonrelevant properties of model$_2$ provide the neutral analogies in the sense that they can be changed around to provide changing sets of positive analogies between model$_2$ and model$_1$. (We assume for simplicity that $\{R\}$ is a proper subset of $\{R'\}$.) In the billiard ball model of a gas we change the model$_2$ to one of 'sticky' billiard balls by introducing forces between the balls—this can now be imported into the gas model to provide a new variant for our theoretical model of gases. The rôle of model$_2$ is to provide a reservoir of neutral analogies, nonrelevant properties of model$_2$, which can be used to provide a sequence of model$_1$'s to deal with anomalies and provide novel predictions as the research programme progresses. Notice how our account of neutral analogies differs from Hesse's. For her neutral analogies are fixed properties of model$_2$ which may be converted into positive analogies in order to develop a sequence of theories But it is the flexibility in the specification of model$_2$, the balls can be sticky or non-sticky, that provides the heuristic for the research programme based on model$_2$.

specify a theory T_n, which is 'approximately' the same as T, for which y_n is an exact solution.

In our example y_n is an exact solution of the equation

$$dy/dx - \lambda y + \lambda^n x^{n-1}/(n-1)! = 0$$

which for small λ is approximately the same as the original equation.

In the more general terms of Redhead [1975] an 'approximate' set of solutions defines a new subset of the underlying function space and thereby a new theory T' say.

Of course we must be careful in talking about approximate equations. They may have quite different solutions from the original equations due to instability effects. Conversely, quite different equations may have approximately (or even exactly) the same particular solutions. The advantage of Redhead's [1975] formulation is that all these possibilities are allowed for in a uniform manner. Denoting by Γ the relevant function space and by γ_T the subset of Γ whose members are permitted by the theory T, so T is to be identified with γ_T, then in an approximation individual points in γ_T are moved to new points in Γ, thus occupying a new subset γ_T' which we identify with the approximate theory T'. The approximation in regard to any particular solution is measured by the 'distance' moved in Γ-space by the corresponding point in γ_T. The situation is illustrated schematically in Figure 1.[1]

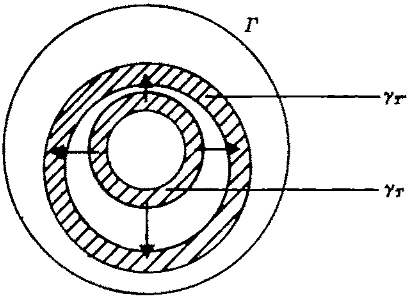


Figure 1

---

[1] Notice how on our account theories are identified with sets of functions whose domain and codomain are part of the specification of the theory. In this way difficulties posed by Miller's [1975] paradoxes concerning the accuracy of predictions, which in the present context translate into ambiguities concerning how we measure a good approximation, are avoided or at any rate side-stepped. Naturally what qualifies as a good approximation depends on what the theory is about and the Millerian charge of essentialism cannot strictly be avoided. However, we are not interested in global measures of the 'distance' between T and T'. In all of our discussions we are concerned with assessing what may be described as local approximations. In this connection we emphasise that the points in Γ include functions and restrictions of functions relating domain and codomain.

Let us suppose that T satisfies the second of Achinstein's criteria, then under what circumstances may we identify T' as a model relative to T or as we may say more shortly a model *for* T? According to the third criterion models are simplified approximations. This may be understood in two senses: Firstly, T' may be mathematically simpler in the intuitive sense that the equations representing T' are simpler. In the above example of the perturbation series the equation for T' was actually more complicated than the equation for T. Secondly, the approximation must be a physical simplification in a sense that can only be established if T' has a direct physical interpretation. How would gases behave if the molecules did not attract one another, what would happen if the nucleons in a nucleus moved in a central field instead of under the influence of their mutual interactions and so on?[1] Mere approximations are introduced in a mathematical context and the question of physical interpretation does not arise.

The question of whether T' possesses an analogy with a more familiar theory T'' is of no fundamental importance. It may be of practical importance in the sense that methods for solving particular problems may already be known for T'' and these can now be imported into the study of T *via* T' and its analogy with T''.

We have seen then that while every impoverishment model is an approximation the converse is by no means true. The approximation involved in the model arises by simplifying in both a physical and mathematical sense the equations governing the theory before solutions are attempted, *i.e.* in treating of models one is concerned with exact solutions to a simplified theory rather than approximate solutions to an exact theory. This is to be contrasted with the claim established above, namely that approximate solutions to exact theories are strictly equivalent to exact solutions to (appropriate) approximate theories. The crucial distinction is between approximation and simplification. The latter notion is a notoriously tricky one to handle with precision, but the attitude of physicists in working with models as opposed to other sorts of approximation is intuitively clear.

In passing we note of course that a model may not be exactly soluble[2] but the idea behind introducing a model is always that its solution is more tractable than for the original theory in its full complexity.

[1] Compare the interesting comment by Del Re in his [1974]: "A simplified treatment based on a model involves replacement of an actual physical system (say a molecule) by a simpler one (the model) which is treated in quite a rigorous way . . . The fact that the approximations are introduced in the model rather than in the mathematical development implies that various aspects of the corresponding treatment can be given a physical interpretation".

[2] For example the Ising model of ferromagnetism was only solved by Onsager by making further approximations, an ideal two-dimensional square lattice.

5 THE USES OF MODELS

So far we have discussed what a model is and the circumstances which give rise to it. We turn now to consider the uses of models and shall take the opportunity to discuss also some putative misuses of the model concept.

The uses of models fall, I think, under three heads:

(*i*) *The Probing of Theories*

In order to work with a complicated theory one first wants to understand the qualitative type of behaviour which the theory can exhibit. Now it is a notorious fact that approximations to a theory, while valid under certain conditions, may give no clue to the way the full theory operates under other conditions even at the grossest qualitative level. By exploring models (what physicists sometimes call tinker toy[1] models) for a theory T we can probe how approximations A_M to the model M, in the style that A_T approximates T, misrepresent M and the true behaviour of M as opposed to A_M can now be used as a guide to how T behaves. Of course the danger in this manoeuvre is that T may behave quite differently from M, which as we have seen is really just another way of representing an approximation to T. But at least the novel behaviour of M can suggest to us how T might behave and lead us to look again at T with this possible behaviour in mind. It is often the case in mathematical physics that we can investigate a solution if we know what sort of thing we are looking for.

Let us take an artificial example to illustrate what we have in mind. Suppose for some phenomenon variables x and y are related by the following differential equation involving a physical parameter λ

$$(1-x)\, dy/dx = \cos \lambda x + y$$

To investigate this 'theory' the physicist might first find an approximate solution in terms of a power series in x. Thus, if we assume the boundary condition $y = 0$ at $x = 0$, the third-order approximation would be

$$y_3 = x + x^2 + (1 - \lambda^2/6)x^3$$

Higher terms in the power series could be calculated but the coefficients become increasingly complicated functions of λ. The approximation y_3 is useful near $x = 0$ but it gives a very misleading picture for larger values of x, in particular near $x = 1$ as we shall see in a moment. Let us now probe the theory in terms of a model M in which we replaced $\cos \lambda x$ by unity—we may call M the small λ model. The corresponding function y_m now satisfies the much simpler equation $(1-x)\, dy_m/dx = 1 + y_m$ which can be written $dy_m/1 + y_m = dx/1 - x$ and integrated immediately to give, with the chosen boundary conditions,

---

[1] Tinker toy models are referred to as study models by Groenwold in his [1961].

$$y_m = x/1-x$$

y_m clearly blows up near $x = 1$ and we can understand the deficiency of our y_3 approximation near $x = 1$ by comparing the corresponding power series expansion for y_m, *viz.*,

$$y_m = x+x^2+x^3+\dots$$

Any finite number of terms in this expansion fails to exhibit the singularity (pole) at $x = 1$. But what about the original function y? From our experience with the small λ model we might look into the possibility of similar behaviour for larger values of λ. This suggests trying as Ansatz for solving the original equation the substitution $y(x) = f(x)/(1-x)$. We easily find that $f(x)$ satisfies the simple equation $df/dx = \cos \lambda x$ which can be integrated immediately with our indicated boundary condition to give now an exact solution for our original problem in the form

$$y = \frac{\sin \lambda x}{\lambda(1-x)}$$

The three functions y, y_m and y_3 are illustrated in Figure 2 taking $\lambda = 3$. Notice that for special values of λ, *viz.* $\lambda = n\pi$, $n = \pm 1, \pm 2 \dots$, the infinite discontinuity at $x = 1$ does not occur for y, so once again y_m misleads us in some respects although it also put us on the road to obtaining the exact solution. So we stress again that the behaviour of M does not necessitate that T behaves in a similar way (it was by recognising this

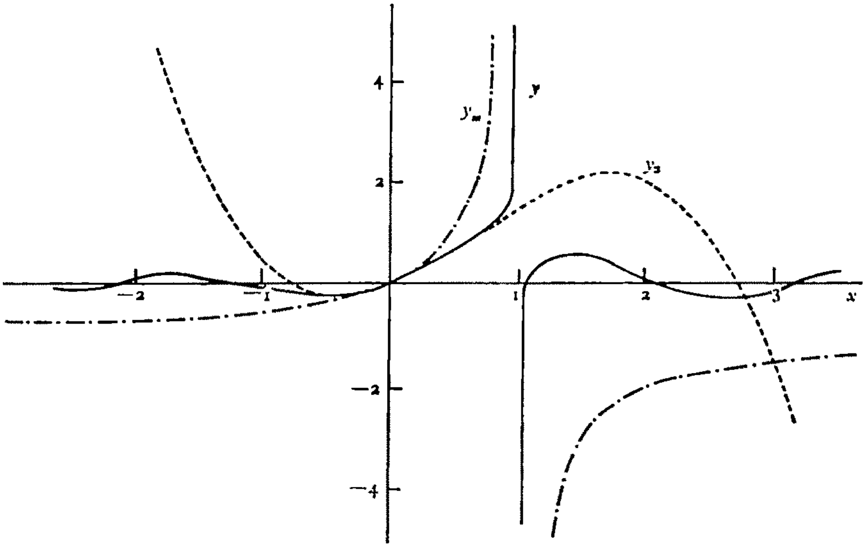


Figure 2

fact in respect of y_3 that we played around with M in the first place). But M has shown how T *could* behave and hence may direct attention to some appropriate aspect of T for careful investigation.

Nevertheless T can be so complicated that we simply do not know whether it behaves like M or not. So what does the physicist do in this situation? He may turn to the second of our three uses of models.

(*ii*) *The Heuristic Rôle of Models*[1]

If a theory T is too complicated to be applied exactly to some group of pheomena a model M may be proposed which is more tractable. This model may have some property, call it C, which serves to explain some aspect of the phenomena. We do not know if the original theory has this property or not, but we may introduce a new theory T' in which the property C is now incorporated as an axiom. The new theory and the old theory may or may not be equivalent, in general we simply do not know, so this is a striking case of a neutral analogy (in Hesse's sense) between a model and a theory being converted into a positive analogy! As an example of this sort of manoeuvre involving approximations rather than models we may cite the development of S-matrix theory as derived from analytic properties of perturbation approximations to field theories. These analytic properties served to predict correlations between observable quantities such as the real part of the forward scattering amplitude and total scattering cross-sections *via* the so-called dispersion relations. But the question remained of whether the full singularity structure of the approximations applied to the unapproximated scattering amplitude. As this question could not be settled by direct calculation Mandelstam in his [1958] imposed this condition as a conjecture which led directly to the development of axiomatic S-matrix theory incorporating the conjecture as an axiom.

Alternatively where the model enriches a theory some particular property may be made model-independent by fiat and fed back as an additional constraint so serving to strengthen the underlying theory. This heuristic rôle of models is illustrated in particle physics by the development of so-called current algebra derived from equal-time commutation relations of specific field models such as the quark model. Gell-Mann describes the situation thus

> We construct a mathematical theory of the strongly interacting particles, which may or may not have anything to do with reality, find suitable algebraic relations

---

[1] The probing of theories as we have seen is also heuristic in the sense of leading (on occasion) to the exact solution of a problem. We are now concerned with heuristic in the sense of discovering new theories rather than solving problems posed by old theories.

that hold in the model, postulate their validity, and then throw away the model. We may compare this process to a method sometimes employed in French cuisine: a piece of pheasant meat is cooked between two slices of veal, which are then discarded.[1]

Again the justification for extracting a particular feature of the model is linked *post hoc* with successful deductions made with its help, for example in the case of current algebra the Adler–Weisberger calculation of the renormalised axial vector coupling constant in β-decay.

We now turn to a third rôle for models.

(*iii*) *The Testing of Theories*

If a theory T is too complicated to work with in producing precise empirical predictions we are faced with what may be called the ***computation gap***. This can only be bridged by means of some approximation which will yield predictions for comparison with experiment. We pursue the case, although this is clearly not essential to the argument, where the approximation is equivalent to a model M. Now suppose M yields a prediction which is contradicted by some basic statement s embodying the result of an observation. We can infer that T is false or that the approximation A used in deriving M is invalid at any rate in respect of phenomena involving s (we use of course the inclusive sense of 'or'). But the mere fact of observing s does not allow us to attribute the refutation of M specifically to T or A. Two posibilities now present themselves. Firstly, we may have independent mathematical reasons for demonstrating that A is a good approximation. In this case we can use s directly as a test of T. But in many cases the relationship between T and M is not so clear cut. If however we have independent reasons for believing that T is true then observing s will show that A is invalid. Similarly in verificationist terms if M (in conjunction with initial conditions) predicts s correctly then s confirms M and if we already believe in T we will conclude that A is valid, *i.e.* that M 'captures' the truth of T in respect of the 'fact' s. This might describe the situation in the quantum mechanics of atomic or molecular systems. The underlying theory, the N-particle Schrödinger equation, is believed because of the success of wave mechanics in explaining the properties of simple systems, the hydrogen atom, the hydrogen molecule and so on. For more complicated systems we use approximations —comparison with experiment is then regarded as a test of the approximation not of the theory. We are doing here a remarkable sort of empirical mathematics, testing mathematical approximations not by examining the situation mathematically but by doing experiments! Of course the fact that the Schrödinger equation gives a good account of simple systems does not as a matter of pure logic entitle us to infer its accuracy in treating complicated systems, so there is an extra inductive step here leading us from simple to complicated systems.

---

[1] See Gell-Mann and Ne'eman [1964], p. 198.

In this situation, where the mathematics is too complicated to test an approximation by placing effective bounds on the errors incurred, another method is often employed that does not depend on assuming T to be true. Suppose two (or more) models are devised which give the same predictions for some observed quantity. This is taken to be evidence that both (all) the approximations are 'adequate'. In order to assess this argument suppose the models $M_1\ M_2 \ldots M_m$ predict values $s_1\ s_2 \ldots s_m$ for some quantity. Let the theory T predict the value s. Then we can define errors $e_1\ e_2 \ldots e_m$ such that $s = s_1+e_1 = s_2+e_2 = \ldots = s_m+e_m$. Now if all the are adequate in respect of the prediction s we may express this situation succinctly as $e_1 = e_2 = \ldots = e_m = 0$. Certainly this condition implies $s_1 = s_2 = \ldots = s_m$. But the reverse implication does not of course follow. Nevertheless, we may argue that the condition $s_1 = s_2 = \ldots = s_m$ supports the claim $e_1 = e_2 = \ldots = e_m = 0$ in the following sense. Let x = prior probability that any one model is adequate. Then, assuming independence, the probability that all m models are adequate is x^m and the probability that all the models are inadequate is $(1-x)^m$. Now m-fold coincidences can only occur if all or none of the m models are adequate, so if ϵ is the probability of coincidence if none of the models is adequate, Bayes's theorem gives

$$P_m = \text{Prob}\left(\begin{matrix}\text{model}\\ \text{prediction is}\\ \text{accurate}\end{matrix}\middle/\begin{matrix}m\text{-fold}\\ \text{coincidence}\end{matrix}\right) = \frac{x^m}{x^m+\epsilon(1-x)^m}$$

$$\epsilon = \frac{9\times 10^{n-1}-1}{(9\times 10^{n-1}-1)^m} \simeq 10^{-n(m-1)}$$

where n is the number of significant digits in the prediction. (The minus one arises because we exclude the coincidence which agrees with the prediction of the theory T.)

Now $P_1 = x/(x+1-x) = x$ as we expect and we are interested in an enhancement ratio

$$\gamma = \frac{P_m}{P_1} = \frac{1}{x+\dfrac{\epsilon}{x^{m-1}}\cdot(1-x)^m}$$

$$\simeq \frac{1}{x+\dfrac{\epsilon}{x^{m-1}}} \text{ if we assume } x \ll 1.$$

We now distinguish two cases

(1) $x^m \gg \epsilon$, giving $\gamma \simeq 1/x \gg 1$

so the posterior probability P_m is very close to unity.

(2) $x^{m-1} \ll \epsilon$, giving $\gamma \simeq x^{m-1}/\epsilon \ll 1$.

In the first case the coincidence has enhanced our expectation that the model prediction is adequate but in the second case the coincidence has actually increased our belief in the *inadequacy* of the prediction. Our analysis shows that the method of coincidence in model predictions is not automatically a sound one for assessing approximations but depends on our having some reasonable prior confidence in the adequacy of the models.[1]

6 FLOATING MODELS

We turn now to consider another sort of model used in testing theories, *viz.*, the floating models recently discussed by Post in his [1974] and [1975]. In this situation we have a model which is disconnected from a fundamental theory T by a computation gap in the sense that we cannot justify mathematically the validity of the approximations being made but which also fails to match experiment with its own (model) predictions. So it is disconnected from the fundamental theory and the empirical facts. In Post's graphic terminology the model 'floats' at both ends. It has, in this sense, no theoretical or empirical support. So of what possible use can it be? Post has parodied the situation with the example of what he calls the Farm Gate Contraction. A farmer investigates the relation between the length of the diagonal strut and the length of the rails and stiles of a farm gate. Although he is familiar with Euclid the derivation of Pythagoras's theorem is utterly beyond his deductive powers. So he invents a model theory, a linear one, in which the lengths are related by

$$l = x + y \tag{1}$$

instead of

$$l = \sqrt{x^2 + y^2} \tag{2}$$

(see Figure 3).

---

[1] As a topical example of the situation we are discussing we may cite the calculations carried out by different methods on the atomic wave functions of bismuth in order to check the apparent refutation of popular unified theories of weak interactions by experimental groups at Oxford and Washington. See Baird *et al.* [1976].

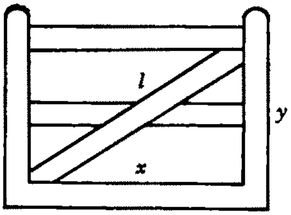


Figure 3

Now (1) has many properties analogous to (2)—for $x = 0$ or $y = 0$ it gives correct values for l and l increases monotonically with x or y in the model as in the correct theory. But detailed measurement shows that (1) is false. So the farmer now introduces a new effect, the Farm Gate Contraction, to explain the mismatch between the predictions of the model and the experimental results. Now of course it is true that any model can be matched with experiment if we allow appropriate modification of the model designed, *ad hoc*, to correct it. So what is the status of the floating model? Indeed has it not misled us into investing a mathematical artifice with a semblance of physical significance (the mythical contraction)? Post has applied these strictures in particular to the Elliott[1] $SU(3)$ model of the nucleus, in which the initially imposed $SU(3)$ symmetry of the wave functions is strongly broken to bring about a match between theoretical and experimental energy levels. But the point about the Elliott model is that the symmetry is not broken in an arbitrary way. The symmetry breaking term in the Hamiltonian is a function of the generators of $SU(3)$ so there is no mixing of wave functions belonging to different irreducible representations of the $SU(3)$ symmetry. In other words physicists only employ floating models if the mismatch can be removed in what is mathematically an intuitively simple way—otherwise any model would be as good as any other if arbitrarily complicated devices for eliminating mismatch were allowed.[2] The adjusted model can then be thought to capture essential features of the underlying theory insofar as this theory is itself believed to be true. By essential features I do not mean that the model approximates to the mechanism of the exact theory in the same way as in the billiard ball model of gases for example, but that some structural features of particular solutions of the exact theory are reproduced approximately in the model. In the case of the Elliott model groups of rotational levels for the exact Hamiltonian in deformed nuclei possess wave functions which support, approximately, irreducible representations of $SU(3)$. So far as the mechanism proposed is concerned, a central field producing harmonic oscillations, and a residual two-body interaction, this is a fiction—all that it tells us is that the indicated properties of the wave functions *could* arise in this way. Thus it is logically but not physically possible to imagine switching off the existing nuclear forces and substituting a central field. So Post is quite right to warn us against investing Farm Gate Contractions with a semblance of reality at the level of explanation, but there is no harm in thinking of the diagonal length *as if* it was caused this way, it certainly gives us the right answer!

---

[1] See Elliott [1958].

[2] I think that physicists are generally aware of what they are doing in using floating models and Post's caricature of the situation is a little unkind to the much-maligned, naive, muddle-headed physicists that philosophers of science like to poke fun at. Nevertheless, there is a natural tendency for experimental physicists to believe in models as being justified mathematically and theoreticians as being proved by experiment.

Another example might help to clarify the issue. Consider the parabolic motion of a projectile under a constant gravitational acceleration. We may consider an 'inertial' model in which we have straight line motion which is then to be corrected by the effect of the gravitational field. Both the model and the correction correspond to physical 'causes' which together give insight into the observed parabolic path.[1] We may say in a different sense from that used above that the inertial model floats only at one end (the observational), since there are now physical as opposed to mathematical grounds for introducing the model. Now consider a 'constant velocity' model where the effect of gravitation is to impart not a uniform acceleration but a uniform velocity. The motion according to the model is again a straight line and the mismatch is just as simple (the difference between a straight line and a parabola) as for the inertial model. So this doubly floating model is acceptable, although both the components in the total explanation are essentially fictions.

---

[1] The question of physical significance of component parts of an explanation is sometimes linked to linearity. In a linear network of cause and effect the total effect when several causes act cooperatively is in a natural sense the sum of the effects when each cause acts separately. Schematically, denoting causes by C and effects by E, $C_1 \rightarrow E_1$, $C_2 \rightarrow E_2$ and $C = C_1+C_2 \rightarrow E = E_1+E_2$. In a nonlinear problem $C = C_1+C_2 \rightarrow E \neq E_1+E_2$. However, suppose we write $E = OC$ where O is some nonsingular nonlinear operator then we can write $E = O(O^{-1}E_1+O^{-1}E_2) = E_1+f(E_1,E_2)$ where $f(E_1, E_2) = O(O^{-1}E_1+O^{-1}E_2)-E_1$. Similarly, $E = E_2+g(E_1, E_2)$ where $g(E_1, E_2) = O(O^{-1}E_1+O^{-1}E_2)-E_2$. These two representations taken together show E_1 and E_2 as component parts of E. The point about the linear case is that the correction terms $f(E_1,E_2)$, $g(E_1, E_2)$ do not now depend on both arguments. It is not clear how this circumstance engenders the 'reality' of E_1 and E_2 as components of the total effect. A simple example is the one-dimensional motion of a particle under the action of two harmonic oscillator potentials associated with different angular frequencies w_1 and w_2. Consider the solution $x = \cos\sqrt{w_1^2+w_2^2}\cdot t \neq \cos w_1 t+\cos w_2 t$, but both potentials are equally 'real' causes of the total effect, the motion of a particle released from rest at unit distance from the origin. I am grateful to one of my students, Douglas Stone, for helpful discussion on the subject of this footnote.

7 CONCLUSION

Let us summarise our conclusions concerning the justification (or lack of justification) for the uses made by physicists of theoretical models.

So far as the probing of theories is concerned the discussion of models employs an argument from analogy—because the model and the theory agree in a number of respects it is argued that they may agree in other, problematic, respects. Such arguments can never of course be regarded as conclusive. At best they enhance the probability of the conclusion and hence may justify the physicist in pursuing the possibility that the model behaviour is also exhibited by the full theory. But physicists should always be aware of the potential fallibility in this mode of reasoning.

In regard to the heuristic rôle of models the technique of generating new theories by extracting certain features of models which are then introduced as axiomatic for the new theory bypasses the difficulties referred to in the probing of theories, since we do not require that the new theory is just a reformulation of the old theory. Now this method is really to be justified *post hoc* by appraising the resulting theory in terms of empirical adequacy, the generation of novel predictions, simplicity of formulation and so on. There is no guarantee that theories generated from models in the way we have described will possess these desirable features to any greater degree than theories generated by heuristic procedures which are not related to models. Since the model is discarded once the new theory has been arrived at we may question what exactly is the rôle of the model in the heuristic process. The thinking here at the intuitive level is that the model has some properties which are to be taken seriously in the sense of being candidates for part of the structure of a new theory. But no logical basis can be given for distinguishing such properties. There does appear at this point to be an irreducibly intuitive element in the practice of physics.

Finally, for testing theories impoverishment models and approximations have been shown to be essential in many cases in view of the computation gap. Nevertheless an unfortunate ambiguity results in attributing refutation of the model predictions to the underlying theory or the approximation procedures. This is a characteristic impasse for example of the particle physics of the strongly-interacting hadrons, for which no soluble elementary phenomena present themselves to provide a direct test of any proposed dynamical theory. The logical basis for justifying the adequacy of models in terms of identical predictions from several different models has been shown to depend critically on one's prior confidence in the adequacy of the models, while in the case of floating models we have established that the physicist's use of these devices is not as arbitrary as is sometimes suggested.

We conclude with a few general remarks concerning the present status of theoretical models. It is tempting to extend the use of the term model to cover any set of simplified and approximate assumptions about some object or system proposed in the context of some underlying theory. Achinstein's second criterion, that models attribute an inner structure, would then be dropped. Certainly modern usage in terms of mathematical models[1] as simplified or idealised representations of physical theories would appear to countenance this. Science has been described as the art of the soluble.[2] Since theories in general are not soluble this can really be translated as science is the art of modelling in the extended sense we have proposed. If we always tried to solve every problem with absolute accuracy, and neglecting no 'accidental' aspects, science could never get started.[3] Science depends on the possibility of ignoring accidents, of isolating certain key features in a situation. These are captured by models, although in the very act of idealisation and approximation we convince ourselves that the model is indeed false. It is moreover in respect of modelling that the imaginative and intuitive element in theoretical physics is most clearly seen so far as normal science in the Kuhnian sense is concerned. Modelling is certainly an art, involving a number of logical gaps which we have tried to indicate in this paper, but even as practised in modern theoretical physics it is not perhaps quite the black art that some philosophers of science would have us believe.

Chelsea College

REFERENCES

Achinstein, P. [1968]: *Concepts of Science: A Philosophical Analysis.*

Andrews, J. G. and McLone, R. R. [1976] (*eds.*): *Mathematical Modelling.*

Baird, P. E. G., Brimicombe, M. W. S. M., Roberts, G. J., Sandars, P. G. H., Stacey, D. N., Fortson, F. N., Lewis, L. I., Lindahl, E. G. and Soreide, D. C. [1976]: 'Search for Parity Non-Conserving Optical Rotation in Atomic Bismuth', *Nature*, **264**, pp. 528–9.

Black, M. [1962]: *Models and Metaphors: Studies in Language and Philosophy.*

Campbell, N. R. [1957]: *Foundations of Science.*

Del Re, G. [1974]: 'Current Problems and Perspectives in the MO-LCAO Theory of Molecules', in Per-Olov Löwdin (*ed.*): *Advances in Quantum Chemistry*, **8**, pp. 95–136.

---

[1] We are not concerned here with the notion we defined for our own purposes in section **3** above. For a discussion of mathematical modelling in the present sense see Andrews and McLone [1976].

[2] Medawar [1967].

[3] Koertge in her [1977] has considered the contrast between Galileo and Guidobaldo del Monte in their attitude to accidents. If Guidobaldo's 'conscientious' approach had predominated science would have stagnated.

Elliott, J. P. [1958]: 'Collective Motions in the Nuclear Shell Model. I. Classification Schemes for States of Mixed Configurations', *Proceedings of the Royal Society of London*, Series A, **245**, pp. 128–45.

Gell-Mann, M. and Ne'eman, Y. [1964]: *The Eightfold Way.*

Groenewold, H. J. [1961]: 'The Model in Physics', in H. Freudenthal (*ed.*): *The Concept and the Rôle of the Model in Mathematics and Natural and Social Sciences*, pp. 98–103.

Hesse, M. B. [1963]: *Models and Analogies in Science.*

Hesse, M. B. [1965]: 'The Explanatory Function of Metaphor', in Y. Bar-Hillel (*ed.*): *Logic, Methodology and Philosophy of Science.* Proceedings of the 1964 International Congress, pp. 249–59.

Hesse, M. B. [1974]: *The Structure of Scientific Inference.*

Koertge, N. [1977]: 'Galileo and the Problem of Accidents', *Journal of the History of Ideas*, **38**, pp. 389-408.

Leatherdale, W. H. [1974]: *The Rôle of Analogy, Model and Metaphor in Science.*

Mandelstam, S. [1958]: 'Determination of the Pion-Nucleon Scattering Amplitude from Dispersion Relations and Unitarity. General Theory', *Physical Review*, **112**, pp. 1344–60.

Medawar, P. B. [1967]: *The Art of the Soluble.*

Miller, D. [1975]: 'The Accuracy of Predictions', *Synthese*, **30**, pp. 159–91.

Post, H. R. [1974]: 'Against Ideologies', Inaugural Lecture, Chelsea College, University of London.

Post, H. R. [1975]: 'The Misuse of Models', paper read at the Boston Colloquium.

Redhead, M. L. G. [1975]: 'Symmetry in Intertheory Relations', *Synthese*, **32**, pp. 77–112.

Brit. J. Phil. Sci. 31 (1980), 165–173 *Printed in Great Britain*

On the Interpretation and Transitivity of Non-standard Synchronisms

by R. FRANCIS

Introduction

1 *The Interpretation of Non-Standard Synchrony*

2 *Direction-dependent Synchrony*

3 *Time-dependent Synchrony*

4 *PCD and Transitivity of Synchrony*

Appendix

INTRODUCTION

In the first section I discuss the interpretation of non-standard synchrony, and state the Principle of Committed Directionality (*PCD*) which is assumed in the balance of the paper. Section 2 provides a derivation of the general expression for ϵ in three spatial dimensions, which agrees with one previously published for two spatial dimensions in the special case that ϵ is symmetric about the x-axis. In an appendix, I illustrate the use of this expression by offering another proof of the intransitivity of non-standard synchronisms amongst non-collinear points. In the third section I consider synchronisations in which ϵ is a function of time. The final section examines the claims of Quinn that collinear synchrony can be intransitive and that there is a method of setting clocks which results in non-standard synchrony among non-collinear points being transitive, and finds that they fail, unless *PCD* is rejected.

1 THE INTERPRETATION OF NON-STANDARD SYNCHRONY

The thesis of the conventionality of simultaneity in the special theory of relativity (*STR*) relies on the observation that postulates concerning the one-way speed of light in a given direction are empirically equivalent, subject to the restriction that the resulting round-trip speed of light is invariably c. This means, for example, that the size of the Lorentz contraction and time dilatation effects are a function of directed spatial orientation, as well as of round-trip speed.[1]

---

[1] *Cf.* Winnie [1970].

Received 12 *February* 1979

Debate concerning non-standard synchrony arose in the context of the causal theory of time,[1] of which it is an essential feature that the causal order of events is independent of any conventions adopted. This permits the assumption of any postulate concerning the one-way speed of light (in a particular direction) satisfying the round-trip principle, provided that if an event e_1 causally precedes an event e_2 under one such postulate, e_1 causally precedes e_2 under any other. In case e_1, e_2 are lightlike connected, this requirement, upon imposition of the round-trip principle, results in the following expressions for the one-way speeds of light in one dimension:[2]

$$c_{\pm} = \frac{c}{2\epsilon_{\pm}} \tag{1}$$

where

$$\epsilon_{-} = 1 - \epsilon_{+}, \tag{2}$$

in which $0 \leqslant \epsilon_{\pm} \leqslant 1$, $c_{+}(c_{-})$ is the one-way speed of light in the positive (negative) x-direction, and ϵ_{+} (ϵ_{-}) is the value for ϵ in the positive (negative) x-direction.

Here I have assumed implicitly that it is oriented spatial direction for which ϵ is defined. Thus, once a particular value for ϵ governing the one-way speed for signals between any two events with non-spacelike separation is chosen, the same value for ϵ holds for the one-way speeds between any two events in the same spatial orientation. I shall call this assumption the Principle of Committed Directionality (*PCD*).

Suppose that x_1, x_2, x_3 are any three spatially collinear points, where $x_1 < x_2 < x_3$, and that the value $\epsilon = \epsilon_1$ is chosen for signals sent from x_1 to x_2. Then *PCD* requires that one is thereby committed to the value $\epsilon = \epsilon_1$ for signals travelling from x_1 to x_3 or from x_2 to x_3, and to the value $\epsilon = (1-\epsilon_1)$ for signals travelling from x_3 to x_2, from x_3 to x_1, or from x_2 to x_1.

Rejection of *PCD* would permit one to pick a different value for ϵ at each spatial point, even along the same direction. Non-standard synchronisms which satisfy *PCD* already result in anisotropic space; any which violate it result in anisotropic space which is also inhomogeneous. No-one has seriously considered such inhomogeneous non-standard synchronisms,[3] although Reichenbach[4] gave a direction-dependent expression for ϵ_{ϕ} which violates *PCD*, since, as $\epsilon_{\pi+\phi} \not\doteq 1-\epsilon_{\phi}$, equation (2) does not hold. (For criticism of Reichenbach's non-standard synchrony, see Grünbaum [1969][1] and Ellis and Bowman [1967].[2])

---

[1] See, *e.g.*, Reichenbach [1958] and Grünbaum [1973].

[2] *Cf.* Grünbaum [1973], ch. 12.

[3] *Cf.* Ellis and Bowman [1967], pp. 119, 121.

[4] Reichenbach [1958], pp. 162 f.

Indeed, since distant synchronisations are usually formulated in terms of light rays, which define spatial directions in *STR*, it is natural to consider only such synchronisations which satisfy *PCD*. (Observe that standard synchrony satisfies *PCD*, since picking $\epsilon = \frac{1}{2}$ everywhere means that (2) is satisfied along any direction.) Nonetheless, nothing in *STR* appears to prohibit other synchronisms. For example, if ϵ is to be considerd a function not only of spatial direction (θ, ϕ), but also of position (x, y, z), the transit time for light travelling from x_1 to x_2 in a straight line is given by

$$t_{12} = \int_{t_1}^{t_2} \mathrm{d}t = \int_{x_1}^{x_2} \frac{2\epsilon(x)}{c}\,\mathrm{d}x$$

and for light travelling from x_2 to x_1 by

$$t_{21} = \int_{x_2}^{x_1} \frac{2(\epsilon(x)-1)}{c}\,\mathrm{d}x = -\int_{x_1}^{x_2} \frac{2(\epsilon(x)-1)}{c}\,\mathrm{d}x,$$

so that the round-trip time is given by

$$t_{121} = t_{12}+t_{21} = \int_{x_1}^{x_2} \frac{2}{c}\,\mathrm{d}x = (2/c)(x_2-x_1).$$

Hence

$$\frac{2(x_2-x_1)}{t_{121}} = c,$$

so that *PCD*-violating synchronisms can satisfy the round-trip principle.

On the other hand, considerable debate still rages over the acceptability of *any* non-standard synchrony.[3] Should such synchronisms which satisfy *PCD* prove inadmissible, so, *a fortiori*, will those which violate it. In what follows I consider only synchronisms which satisfy *PCD*.

2 DIRECTION-DEPENDENT SYNCHRONY

Suppose that a light ray is emitted from the origin (0, 0, 0) at time $t = 0$, reaches the point $(x, 0, 0)$ at t_1, where it is reflected towards the point $(x, y, 0)$ which it reaches at t_2. Finally, it is reflected back to the origin where it arrives at t_3 (Fig. 1). Define

$$\theta = \sin^{-1}\frac{x}{\sqrt{x^2+y^2}} = \cos^{-1}\frac{x}{\sqrt{x^2+y^2}},$$

and let ϵ_x, ϵ_y be the values for ϵ along the positive x- and y-axes. Then, using equation (1), we have, setting $c = 1$,

---

[1] Grünbaum [1969], pp. 677 f. and p. 853.

[2] *Op. cit.*, pp. 121 f.

[3] See, *e.g.*, Malament [1977] and Francis [1978].

$$t_1 = 2\epsilon_x x \quad \text{and} \quad t_2 - t_1 = 2\epsilon_y y,$$

which yield

$$t_2 = 2\epsilon_x x + 2\epsilon_y y. \tag{3}$$

From the round-trip principle, we have

$$t_3 = x + y + \sqrt{x^2 + y^2},$$

whence, using (3), we obtain

$$t_2 - t_3 = (2\epsilon_x - 1)x + (2\epsilon_y - 1)y - \sqrt{x^2 + y^2}. \tag{4}$$

But, where ϵ_θ is the value for ϵ along the direction θ, we know also from (1) that

$$t_2 - t_3 = 2\sqrt{x^2 + y^2}\,(\epsilon_\theta - 1). \tag{5}$$

Comparison of (4) and (5) produces, upon substitution for sin θ and cos θ,

$$\epsilon_\theta = \tfrac{1}{2} + (\epsilon_x - \tfrac{1}{2}) \cos \theta + (\epsilon_y - \tfrac{1}{2}) \sin \theta. \tag{6}$$

This expression coincides with the following one given by Ellis and Bowman,[1]

$$\epsilon_\theta = \tfrac{1}{2} + (\epsilon_x - \tfrac{1}{2}) \cos \theta,$$

for the special case that $\epsilon_\theta = \epsilon_{-\theta}$, which, as Winnie has remarked,[2] is equivalent to assuming $\epsilon_y = \frac{1}{2}$.

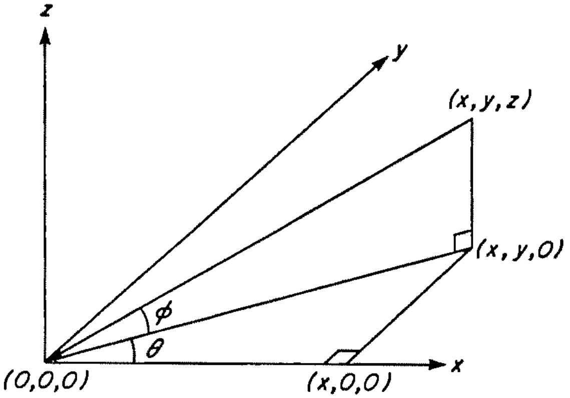


Figure 1

Now suppose that a light ray leaves the origin at $t = 0$, is reflected from $(x, y, 0)$ at t_1, and from (x, y, z) at t_2, returning to the origin at t_3. Define

$$\phi = \cos^{-1} \frac{\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2 + z^2}} = \sin^{-1} \frac{z}{\sqrt{x^2 + y^2 + z^2}}.$$

---

[1] *Op. cit.*, p. 120.

[2] *Op. cit.*, p. 223.

Using a procedure identical to that used above, we obtain, after elementary manipulation,

$$\epsilon_{\theta\phi} = \tfrac{1}{2}+(\epsilon_{\theta}-\tfrac{1}{2})\, cos\ \phi+(\epsilon_{z}-\tfrac{1}{2})\, sin\ \phi, \tag{7}$$

where ϵ_z is the value for ϵ along the positive z-axis. Finally, substitution of (6) into (7) produces

$$\epsilon_{\theta\phi} = \tfrac{1}{2}+ [(\epsilon_{x}-\tfrac{1}{2})\, cos\ \theta+(\epsilon_{y}-\tfrac{1}{2})\, sin\ \theta]\, cos\ \phi+(\epsilon_{z} - \tfrac{1}{2})\, sin\ \phi. \tag{8}$$

In the appendix, I illustrate the use of this expression in proving the intransitivity of non-standard synchrony amongst non-collinear points.

3 TIME-DEPENDENT SYNCHRONY

I have considered above the dependence of ϵ on spatial position and direction; in the spirit of Minkowskian relativity one might also consider values of ϵ which vary with temporal position and direction, a possibility Reichenbach mentioned in passing.[1] The non-standard synchronisms examined earlier result in spatial anisotropy and inhomogeneity; time-dependent synchronisms result in temporal anisotropy and inhomogeneity.

The issues are not independent. Since a signal (in one spatial dimension) travelling, say, from left to right in ordinary time becomes one travelling from right to left in reversed time, we have, as a generalisation of (2),

$$\begin{aligned} \epsilon^{\downarrow}_{x_1x_2} &= 1-\epsilon^{\uparrow}_{x_1x_2}, \\ &= \epsilon^{\uparrow}_{x_2x_1} \end{aligned} \tag{2'}$$

where $\epsilon^{\uparrow}_{x_1x_2}$ $(\epsilon^{\downarrow}_{x_1x_2})$ is the value for $\epsilon_{x_1x_2}$ under the ordinary (reversed) time orientation. That is, the combined operations of parity and time reversal preserve the value of ϵ for a particular signal.

A still more general kind of non-standard synchrony involves values of ϵ which, along a given direction, are (integrable) functions of time, as well as of position. For such synchronisms, in a given physical situation, ϵ must be averaged over time and space along the signal path. For example,

$$\epsilon = \frac{1}{XT}\int_0^X \int_0^T \epsilon(x, t)\, \mathrm{d}t\, \mathrm{d}x$$

is the value for ϵ holding for the straight space-time path with end-points (0, 0, 0, 0), (T, X, 0, 0).

At this point, an objection to temporally inhomogeneous synchronisms might be raised along the following lines. The conventionality of distant simultaneity is due to the freedom *STR* permits in setting distant clocks with respect to a corresponding freedom of choice for the one-way speeds.

---

[1] *Op. cit.*, p. 168.

In allowing ϵ (and hence the one-way speeds) to be time-dependent, is one not violating a sort of 'vicious circle' principle?

The answer to this brings out an important assumption underlying the present paper. I am taking a standard ($\epsilon = \frac{1}{2}$) space-time co-ordinatisation as fixed at the outset (in the given inertial frame). The occurrences of 'x', 't' in $\epsilon(x, t)$ refer to this co-ordinatisation. From this viewpoint, any non-standard synchronism is seen as picking out a set of space-time points with *different* time co-ordinates as simultaneous.[1]

4 PCD AND TRANSITIVITY OF SYNCHRONY

The essential features of Quinn's discussion [1974] are as follows:

(*i*) There is a procedure for synchronising three collinear clocks at spatial points A, B, C such that

$$\epsilon_{AB} = \epsilon_{BC} = \epsilon_K \neq \tfrac{1}{2} \tag{9}$$

and

$$\epsilon_{AC} \neq \epsilon_K \tag{10}$$

Hence, he argues, Grünbaum's proof[2] that non-standard synchronisms are intransitive is extendible to the case of collinear clocks, provided Grünbaum's 'two-ray' procedure is employed.

(*ii*) If, on the other hand, Quinn's 'one-ray' procedure is used, non-standard synchronisms are transitive for collinear clocks. Quinn concludes that statements concerning the transitivity of synchronisations must be relativised to the method followed in setting the clocks.

(*iii*) Quinn further claims that these results are generalisable for non-collinear clocks, so that, from (*ii*), there is a procedure for setting clocks, collinear or not, which results in even non-standard synchronisms being transitive.

I shall deal with these points in turn. In the third section of his paper, Quinn considers three collinear clocks A, B, C, where A is between B and C. It follows immediately from *PCD* that

$$\epsilon_{BA} = \epsilon_{AC} = \epsilon_{BC} \tag{11}$$

and

$$\epsilon_{AB} = \epsilon_{CA} = (1 - \epsilon_{BA}). \tag{12}$$

Hence

$$\epsilon_{AB} = (1 - \epsilon_{BC}). \tag{13}$$

---

[1] For details, see Francis [1978].

[2] Grünbaum [1969], pp. 674 ff.

Comparison of the first half of Quinn's premiss (9) with (13) produces

$$\epsilon_{AB} = (1-\epsilon_{AB}),$$

so that $\epsilon_{AB} = \frac{1}{2}$, which contradicts the second half of (9). I conclude that the premiss (9) and conclusion (10) of Quinn's argument (*i*) are inconsistent with (13) and hence with *PCD*.

In his second section, Quinn maintains that Grünbaum's proof of the intransitivity of non-standard synchronisms depends crucially on his use of a two-ray procedure for setting clocks, and that synchronisation using a single ray results in all non-standard synchronisms being transitive. Now, if my treatment above of synchrony between collinear points is correct, *all* synchronisms, standard or not, are trivially transitive for such points. (It should be pointed out that Grünbaum's proof plainly deals only with non-collinear points.)[1] Roxburgh[2] and Leplin[3] have observed that Quinn's procedure is distinguished not by the number of rays used, but rather by his requirement that all clocks are to be set and compared using the same light path.

For example, if clocks at *B* and *C* are synchronised with one at *A* by sending a light ray from *A* to *B* and thence to *C*, all light rays used to check synchrony with respect to *A* of any other clock must also pass through *B*. This means that for Quinn (and also for Roxburgh and Leplin), ϵ_{AC} is not the value for ϵ along the line passing through *A* and *C* (unless *A*, *B*, *C* are collinear), but instead a path-length average of ϵ_{AB} and ϵ_{BC}. If such path-averaged values for ϵ are to be used in determining transitivity, then all synchronisms do indeed turn out to be transitive. But such a conception of transitivity violates *PCD*; in fact it is just this principle which distinguishes transitivity amongst collinear points from transitivity amongst non-collinear points. Consequently, unless Quinn rejects *PCD*, his claim (*iii*) fails.

APPENDIX

As an illustration in the application of equation (8), I use it to derive Grünbaum's result[4] that all non-standard synchronisms (amongst non-collinear points) are intransitive.

Consider three points *A*, *B*, *C*. Without loss of generality we may assume that *A*, *B*, *C* have co-ordinates (0, 0, 0), $(x_1, y_1, 0)$, $(x_2, 0, 0)$

---

[1] *Ibid.*, p. 675.

[2] Roxburgh [1975].

[3] Leplin [1976].

[4] Grünbaum [1969], pp. 674–6.

respectively, where $x_2 > x_1 > 0$ (Fig. 2). In this case, since $\phi = 0$, equation (8) reduces to equation (6). Upon defining

$$\theta_1 = \sin^{-1}\frac{y_1}{\sqrt{x_1^2+y_1^2}} = \cos^{-1}\frac{x_1}{\sqrt{x_1^2+y_1^2}}$$

and

$$\theta_2 = \sin^{-1}\frac{y_1}{\sqrt{(x_2-x_1)^2+y_1^2}} = \cos^{-1}\frac{x_2-x_1,}{\sqrt{(x_2-x_1)^2+y_1^2}}$$

we have

$$\epsilon_{AB} = \tfrac{1}{2}+(\epsilon_y-\tfrac{1}{2})\sin\theta_1+(\epsilon_x-\tfrac{1}{2})\cos\theta_1, \tag{14}$$

$$\epsilon_{BC} = \tfrac{1}{2}+(\epsilon_y-\tfrac{1}{2})\sin(-\theta_2)+(\epsilon_x-\tfrac{1}{2})\cos(-\theta_2), \tag{15}$$

and

$$\epsilon_{AC} = \epsilon_x. \tag{16}$$

A synchronisation is intransitive if $\epsilon_{AB} = \epsilon_{BC}$ implies $\epsilon_{AC} \neq \epsilon_{AB}$.

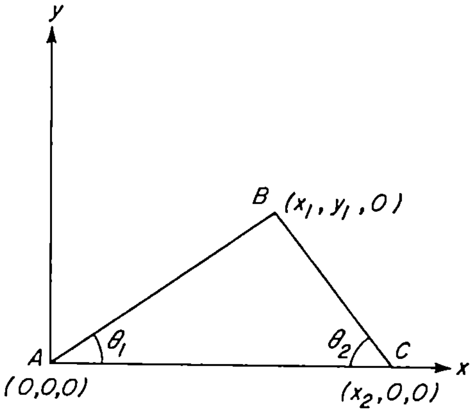


Figure 2

Here, two remarks are in order. First, in standard synchrony, $\epsilon_x = \epsilon_y = \frac{1}{2}$, so that $\epsilon_{AB} = \epsilon_{BC} = \epsilon_{AC} = \frac{1}{2}$. Hence all standard synchronisations are transitive. Second, if A, B, C are collinear, so that θ_1 and θ_2 vanish, then $\epsilon_{AB} = \epsilon_{BC} = \epsilon_{AC}$, no matter what value for ϵ_{AB} is chosen. Thus all synchronisations amongst collinear points are transitive.

To prove the intransitivity of non-standard synchronisations amongst non-collinear points, suppose that θ_1, θ_2, $\neq 0, \pi$, and $\epsilon_{AB} = \epsilon_{BC} \neq \frac{1}{2}$. We must show that $\epsilon_{AC} \neq \epsilon_{AB}$. If $\epsilon_{AB} = \epsilon_{BC}$, substitution from (14) and (15) yields

$$(\epsilon_y-\tfrac{1}{2}) = (\epsilon_x-\tfrac{1}{2})\frac{\cos\theta_2-\cos\theta_1}{\sin\theta_1+\sin\theta_2}, \tag{16'}$$

since $\theta_1, \theta_2 \neq 0, \pi$. Now substitute (16′) into (14):

$$(\epsilon_{AB}-\tfrac{1}{2}) = (\epsilon_x-\tfrac{1}{2})\left[\frac{\cos\theta_2-\cos\theta_1}{\sin\theta_1+\sin\theta_2}\sin\theta_1+\cos\theta_1\right]. \tag{17}$$

Suppose $\epsilon_{AB} = \epsilon_x = \epsilon_{AC}$. Then from equation (17), either $\epsilon_{AB} = \epsilon_x = \frac{1}{2}$, which contradicts the assumption that the synchronisation is non-standard, or

$$\left[\frac{\cos\theta_2-\cos\theta_1}{\sin\theta_1+\sin\theta_2}\sin\theta_1+\cos\theta_1\right] = 1. \tag{18}$$

But use of trigonometric identities shows that (18) is unsatisfiable under the assumed constraint that $\theta_1, \theta_2 \neq 0, \pi$.

Hence $\epsilon_{AB} \neq \epsilon_x = \epsilon_{AC}$, and non-standard synchronisations amongst non-collinear points are intransitive.[1]

University of Toronto

REFERENCES

Ellis, B. and Bowman, P. [1967]: 'Conventionality in Distant Simultaneity', *Philosophy of Science*, **34**, pp. 116–36.

Francis, R. [1978]: 'Malament and Distant Simultaneity: Conventionality Regained' (forthcoming).

Grünbaum, A. [1969]: 'Simultaneity by Slow Clock Transport in the Special Theory of Relativity', *Philosophy of Science*, **36**, pp. 5–43; reprinted in Grünbaum [1973], pp. 670–708, to which page references are made.

Grünbaum, A. [1973]: *Philosophical Problems of Space and Time* (Second, enlarged edition).

Leplin, J. [1976]: 'Synchronisation Rules and Transitivity', *The British Journal for the Philosophy of Science*, **27**, pp. 399–402.

Malament, D. [1977]: 'Causal Theories of Time and the Conventionality of Simultaneity', *Noûs*, **11**, pp. 293–300.

Quinn, P. L. [1974]: 'The Transitivity of Non-Standard Synchronisms', *The British Journal for the Philosophy of Science*, **25**, pp. 78–82.

Reichenbach, H. [1958]: *The Philosophy of Space and Time*.

Roxburgh, I. [1975]: 'The Intransitivity of Non—Standard Synchronisms', *The British Journal for the Philosophy of Science*, **26**, pp. 47–9.

Winnie, J. [1970]: 'Special Relativity without One-Way Velocity Assumptions', *Philosophy of Science*, **37**, part I: pp. 81–99; part II: pp. 223–38.

---

[1] [*Added in Press*]: Since the writing of this note, Quinn (private communication) has informed me that he rejects *PCD*. In view of this, section 4 should be taken, not as demonstrating an error in Quinn's discussion, but rather as educing the inconsistency of his synchronisation procedures with *PCD*. A consequence of his methods of synchronisation, then, is that the speed of light may vary continuously in space, subject only to the round-trip principle. Further, in the absence of *PCD*, the concept of transitivity is significantly weakened. I am grateful to Professor Quinn for his comments.

Brit. J. Phil. Sci. 31 (1980), 175–180 *Printed in Great Britain*

Discussion

SOME COMMENTS ON THIRD WORLD EPISTEMOLOGY

1 Sir Karl Popper's striking hypothesis about a third world of objective knowledge deserves a more careful scrutiny than philosophers have hitherto given it. The hypothesis asserts an ontological thesis which can easily be adapted to fit certain non-Popperian methodologies, such as inductivism. So the hard core of the hypothesis is not hit by standard criticisms of Popperian methodology, and needs separate consideration. If it survives criticism, it will stand as a rather powerful theory about the nature of human culture: if it does not, then some reasons will at last have been stated why it can safely be set aside.

2 On Popper's view the world of objective knowledge as he calls it, or more generally the world of 'the objective spirit', is man-made ([1972], p. 147). But it exists to a large extent autonomously, because the moment our critical and creative thinking has produced a theory that theory may itself produce further problems. It follows that when any of these further problems is formulated by a human being, as in Goldbach's conjecture, the problem has been discovered, not made, by us ([1972], pp. 160–1). Indeed a large part of the objective third world of actual and potential theories and books and arguments arises as an unintended by-product of the actually produced books and arguments and of human language ([1972], p. 117). There is thus a constant feed-back between ourselves and our work. It follows that the main aim of all historical understanding is the hypothetical reconstruction of a third-world problem situation. For example, how is Galileo's theory of the tides to be understood? In order to achieve such understanding, says Popper ([1972], pp. 170–5), we have to investigate a fairly complex third-world object. The object consists not just of the standard problem about the cause of tidal motion, but of this together with its then current background of argument about the truth or falsity of Copernican theory, with a certain given theoretical framework of conservation laws.

It follows that the contents of the third world are rather diverse. Some, according to Popper ([1972], p. 109), are just what Frege meant when he talked about the objective contents of human thoughts. Indeed Popper tells us ([1972], p. 157) that 'theories, or propositions, or statements are the most important third-world linguistic entities'. But we are told also ([1972], p. 106) that the third world contains poetic thoughts as well as scientific ones. Indeed it contains works of art, theoretical systems, problem situations, critical arguments, discussion-states, the contents of books and journals, and also scientific conjectures or hypotheses, experiments and their evaluations ([1972], pp. 106–11). Moreover just as the elements of the third world are thus richly diverse in character, so too are the processes in which they participate. There is not only a process of growth whereby a first problem gives way via tentative theorising and error elimination to a second problem ([1972], p. 121). There is also competition and conflict between theories and problems, aspects of conjectures, interpretations and philosophical positions, and there are also comparisons, contrasts, analogies, problem-shifts, and so on ([1972], p. 165). Nor are all the objects in the third world of equal merit, since some objective knowledge is true, some false, some useful and some useless ([1972], p. 115). Indeed, while many theories in the third world are consistent, some—like Frege's *Grundgesetze* —are inconsistent, according to Popper ([1977], p. 57). As Popper ([1972], p. 121) puts it

> All work in science is work directed towards the growth of objective knowledge. We are workers who are adding to the growth of objective knowledge as masons work in a cathedral. Our work is fallible, like all human work. We constantly make mistakes, and there are objective standards of which we may fall short—standards of truth, content, validity, and others.

But the actual pattern of knowledge-growth is a Darwinian, evolutionary one. The fittest hypothesis survives ([1972], pp. 261–4).

Nor should the reader suppose that Popper is just indulging in an enormous metaphor or simile. He makes it quite clear ([1972], p. 106) by the comparison of his hypothesis with Plato's theory of Forms, with Bolzano's theory of propositions in themselves, and with Frege's theory of objective thoughts, that he is ontologising. Indeed he says in so many words that he is philosophically a 'realist' not only in regard to the physical world and the world of states of consciousness, but also in regard to the third world of objective knowledge ([1972], p. 107). His theory is meant to describe how knowledge really grows; it is not meant metaphorically ([1972], p. 261).

Unfortunately, however, despite the length at which Popper has written about his third-world hypothesis, a number of rather important issues remain unclarified.

3 First, there is a problem about inter-communication. If intelligent creatures exist elsewhere in the universe with whom we terrestrial humans will never be able to communicate, the progress of objective knowledge may have gone much further than we shall ever know. If so, no human investigator will ever be able to test whether he has correctly conjectured the objective problem-situation in a particular area of enquiry at a particular time. Indeed exactly the same problem arises in regard to any two cultures in the terrestrial past, if distance, xenophobia, geographical ignorance, or navigational incompetence prevented them from communicating information to one another. It follows that you cannot explain Galileo's theory of the tides, for example, by reference to the objective problem-situation in Galileo's time but only by reference to Galileo's subjective belief about that problem-situation. The explanatory goal of the third-world hypothesis seems unattainable.

One obvious solution of this difficulty would be for Popper to postulate a different world of objective knowledge for every closed group of inter-communicating intelligences. This would certainly deal with the problem across space, as it were. But, in view of the possibility that any such group of intercommunicating intelligences might be the outcome of past mergers or might give rise to future splits, Popper would also have to provide for corresponding processes of merging and splitting in his plurality of worlds of objective knowledge. And if the number of these worlds could thus fluctuate the conception of a third world of objective knowledge would hardly be on all fours with the conception of a physical world or of a world of states of consciousness. It is not now the conception of a totality of entities of a certain ontological category, but comparable rather to such less philosophical conceptions as that of 'the world of ancient China' or 'the world of the remoter galaxies'.

Accordingly, if the ontological unity of the third world is to be maintained, it would be necessary instead to suppose a potentially divided world, wherein perhaps many distinct cathedrals of objective knowledge were in process of mutually dependent erection and were at different stages of growth. But the autonomy of the third world is now severely limited. Its ability to generate creative problems is now restricted by uncrossable frontiers between some of the objects that it contains. And these uncrossable frontiers are not the outcome or the explanation of any intrinsic epistemological factors. They are merely postulated *ad hoc* as the counterparts of certain physical barriers between groups of intelligent creatures.

I do not know which of these three possible responses would seem least objectionable to Popper. But his third-world hypothesis can hardly come to grips with the problem of intercommunication without losing at least some of its initial simplicity and metaphysical glamour.

4 There is a worse difficulty, however. 'Every scientific theory', says Popper ([1977], p. 140), 'has, objectively, a huge set of important consequences, whether or not these have as yet been discovered'. That is how the problem about Goldbach's conjecture came to exist in the third world long before it was actually discovered. And the third world also contains inconsistent theories, like Frege's *Grundgesetze*. But according to classical logic a contradiction implies anything whatever. So whatever proposition you care to mention is already derivable within some third-world theory. It looks as though the masons who are building the grand cathedral of objective knowledge have not only completed it but also rather disastrously over-fulfilled their norms.

How can this difficulty in Popper's hypothesis be circumvented? It is tempting to suppose that, in relation to every theory contained by it, the third world also contains the fact that the theory is consistent or inconsistent, as the case may be. Indeed Popper himself ([1977], p. 57) writes of 'the objective hitherto unnoticed fact, belonging to World 3, that there was an inconsistency in Frege's theory'. After all, if the third world is attributed the capacity, as it were, to generate logical consequences, the capacity to detect inconsistencies might just as well be attributed to it. Then in accordance with such detections inconsistent theories can be consigned to some special category of demerit within the third world. But has the third world a solution for the decision-problem in regard to polyadic quantification-theory? No, for presumably Church's theorem is to be interpreted as a proof that the third world cannot contain a solution to this problem. It follows that the third world can hardly be supposed uniformly capable of including, whenever a new theory is fed into it, the fact that it is consistent or inconsistent, as the case may be; or we should be attributing to the third world a capacity that is demonstrably impossible.

It may seem preferable therefore to suppose that the third-world operates with a relevance logic or with some other criterion of logical consequence that filters out any logical consequences that are due purely to the inconsistency of their premisses. But any such logic is a topic of further controversy, and it would be interesting to know how Popper himself would propose to deal with the logical problem here, without bringing in unwanted paradoxes.

5 A further difficulty arises as a result of Popper's claim that the third world contains all the consequences of any scientific theory, whether or not these have as yet been discovered. For it seems that the third world may thus contain not only the problem for which Goldbach conjectured his solution but also the correct solution. If it is derivable from the postulates of number theory that every even number is the sum of two primes, or that some even number is not such a sum, then one or other of these two theorems ought already to be in the third world. It got there just as soon as the postulates did. Of course, the proof may not be there yet, and perhaps will never be. And perhaps, after all, Goldbach's theorem is neither provable nor disprovable. But clearly Popper's hypothesis entails that the third world contains an enormous amount of mathematical knowledge—in the form of provable but unproved theorems—which is as yet unknown to any human being. It follows inevitably that the growth of such knowledge cannot be a third world process. In this respect at least the growth of human knowledge can only be represented by a growth of subjective awareness, as human beings come gradually to discover provabilities that are already (according to Popper) part of objective knowledge. Correspondingly the problem-situation at any particular time has to be viewed in subjective terms, not in objective ones, since it is only in human minds, not in the third world, that any of the relevant problems are as yet unsolved.

Nor is this just a difficulty for the Popperian epistemology or mathematics: it affects the history of natural science also. The apparent discrepancy between Newton's theory and Flamsteed's observations, in regard to the motion of the lunar apse, created an important problem-situation in early eighteenth-century mechanics (*cf.* Cajori [1962], pp. 648–50). Even Clairaut at first thought that the motion of the lunar apse was wholly inexplicable by Newton's laws, and started to formulate a new hypothesis regarding gravitation. But eventually he was able to carry out calculations showing that the observational results were indeed predictable from Newton's laws. Now according to Popper's hypothesis, if Newton's laws, plus the relevant statements of initial conditions, were already in the third world when Clairaut began his investigations, then so too was the fact that the appropriate motion of the lunar apse is derivable from these premisses. So, if the problem-situation that Clairaut encountered had really been determined by the third-world situation, he would never have wasted any time on trying to develop a rival hypothesis to Newton's and would instead have concentrated on his new calculations from the start. The only way for a historian to understand what actually happened at first is by reference to Clairaut's subjective but mistaken belief about the problem-situation.

It would be open to Popper to exclude this kind of difficulty in his third-world hypothesis by revising his view so as to deny that the logical consequences of a theory enter the third world before some human scientist discerns them. But then the autonomy of the third world would be gravely diminished. The inconsistency in Frege's *Grundgesetze*, for example, would not have entered the third world until Russell discovered it. So it wouldn't have been the objective autonomy of the third world that threw up the inconsistency, as it were, but the subjective creativeness of Frege's mind. Equally Goldbach would have to be given the credit (which he surely deserves) for having posed the question that his conjecture was designed to answer. All that would have pre-existed Goldbach's question was the terminology in which it was formulated, along with some other (presumably suggestive) questions about similar topics.

6 It might be urged, on behalf of some supposed hard core of Popper's hypothesis, that even when one speaks of, say, Clairaut's mistaken belief about the problem-situation, one assumes the existence of an objective problem-situation for Clairaut to misconceive. Therefore, it might be said, a third world has still to be postulated in which such objective problem-situations enjoy their ontologically distinctive mode of existence. Maybe, but there is certainly a difficulty in regard to the epistemology of beliefs or theories about the contemporary problem-situation, like those beliefs or theories attributable on Popper's hypothesis to Galileo or Clairaut. These beliefs or theories seem themselves to be indispensable components in the growth of human knowledge. But they are second-order components as it were: they issue in judgments or conjectures about other judgments or conjectures. For example, suppose in a particular field of enquiry theory T is the best theory so far proposed. Presumably Popper's third world contains some objective fact to this effect. Suppose too that only one investigator has so far considered T, and *his* theory (T') is that T is not the best theory so far proposed. It seems undeniable then that T' deserves a place in the third world alongside T, and indeed that T' is itself the best theory so far proposed about T. No doubt T' is false, but T may well be false also, so that there is nothing to choose between T and T' on this score. Accordingly it looks as though third-world declarations of theory-merit are potentially incoherent. Of course, this incoherence could easily be avoided by adding some special postulate to Popper's hypothesis: second-order theories would just have to be banned from consideration in connection with the status of first-order ones. But that is as if second-order theories contributed to the growth of knowledge in a quite separate world—a fourth world of objective meta-knowledge—and there is an obvious danger here of a regress into worlds without end. Alternatively we could say that only correct second-order theories are admissible to the third world. But that seems out of keeping with Popper's general principle that neither the falsehood of a theory nor its inconsistency is a sufficient reason to suppose it inadmissible. Moreover, the more features that have to be postulated *ad hoc*, in order to circumvent particular difficulties in Popper's third-world hypothesis, the less we may feel inclined to indulge in the ontological extravagance that this hypothesis intrinsically involves.

7 Finally there is a serious difficulty about Popper's claim ([1972], p. 261) that his epistemology is 'a largely Darwinian theory of the growth of Knowledge'. On this view theories compete for survival and the fittest survives. But which are the species and which are the species-members in this epistemological Darwinianism? If theories are cast in the role of species, then presumably the species-members are the representations of these theories in the minds and writings of individual scientists. But if that is so then the real struggle for survival takes place in the first and second worlds, not in the third. And yet a refuted theory is not automatically deleted from all the books in which it was printed. Many tokens of it remain, even when its more successful rivals enjoy much greater popularity. It follows that competition between theories cannot be envisaged as a Darwinian struggle in which theories operate as species and theory-representations as species-members. Are then the theories themselves to be conceived of as species-members, competing for individual survival in the third world? But if so they seem for the most part rather ill-suited to constitute species. They serve their intended purposes only in virtue of being distinctively tailor-made for these and therefore different from one another. In sum, one is inevitably driven to the conclusion that any talk about Darwinian evolution in Popper's third world is either false or just metaphorical, contrary to Popper's own claim ([1972], p. 261). The complaint here is not that developments in Popper's third world are not the result of genetic mutations or physiological adaptations. How could they be? The trouble is rather that the central feature of the evolutionary process has no counterpart in the third world. Speciation does not play any important role in third-world development. Yet it was what Darwin wrote his great book about.

8 It remains to be seen whether Popper's hypothesis can be adapted to meet the above-mentioned difficulties in a more plausible fashion than any that I have been able to propose. If no such adaptation can be developed, the hypothesis seems scarcely worth preserving.

L. JONATHAN COHEN
The Queen's College, Oxford

REFERENCES

CAJORI, F. [1962]: 'Appendix' in Cajori (*ed.*) *Sir Isaac Newton's Mathematical Principals of Natural Philosophy and his System of the World.* Berkeley: University of California Press.

POPPER, K. R. [1972]: *Objective Knowledge: An Evolutionary Approach.* Oxford: University Press.

POPPER, K. R. and ECCLES, J. C. [1977]: *The Self and Its Brain.* Berlin: Springer-Verlag.

Brit. J. Phil. Sci. 31 (1980), 181–192 *Printed in Great Britain*

Review Article

THE HALT AND THE BLIND: PHILOSOPHY AND HISTORY OF SCIENCE*

The opening chapter of *Method and Appraisal in the Physical Sciences* is Imre Lakatos's 'History of Science and its Rational Reconstructions', a paper first presented to the Philosophy of Science Association in October 1970 and quickly printed in *Boston Studies*. I was then the commentator (though I saw Lakatos once more, this was the occasion of our last intellectual exchange), and my response included the following sentence: 'What Lakatos conceives as history is not history at all but philosophy fabricating examples'.[1] *Method and Appraisal* can be read as in part a test of that comment, for its editor, Colin Howson, places five historical case studies between Lakatos's opening essay and a concluding critique by Paul Feyerabend. In the event these historical chapters are a considerable success, and I have not been altogether able to set aside the question how this could be so. Parts of the answer seem worth recording and will appear here and there below.

Lakatos opens his paper with sketches of four current methodological positions: inductivism, conventionalism, methodological falsificationism, and his own methodology of research programmes. Summarizing a view developed elsewhere by Agassi, he points out that each of these schools repeatedly makes use of historical examples, but that each tells the story of particular historical episodes in a systematically different way. To explain why this occurs Lakatos draws a sharp line between what he calls 'internal history' and 'external history'. The former is exclusively concerned to provide 'a rational explanation of the growth of objective knowledge', and it is correspondingly selective. 'Philosophy of science [,which] provides normative methodologies in terms of which the historian reconstructs "internal history"', supplies the required principle of selection (p. 1). Two consequences follow: first, because different methodological schools provide different criteria of rationality, internal accounts of scientific development vary from school to school; second, because scientific development is not shaped exclusively by rational forces, internal histories may not, in principle, aim at completeness. Recourse to external history is often needed if one is to explain, for example, why a particular development occurred when and as it did or how a particular group of scientists could fail to recognise an apparent implication of their position. For Lakatos, however, concerns of this sort are secondary. He emphasises that external history is merely the residuum left by internal history and that it is in principle 'irrevelant for the understanding of science' (p. 2). Part 1 of his paper develops these themes and supplies numerous illustrations.

---

* Review of Howson, C. (*ed.*) [1976]: *Method and Appraisal in the Physical Sciences: The Critical Background to Modern Science, 1800–1905*. Cambridge: Cambridge University Press. vii+344pp. £10.50.

[1] See my [1971], p. 143.

In Part 2 Lakatos introduces a series of newer and perhaps more controversial claims. His 'basic idea', he writes, is that 'all methodologies function as historiographical (or meta-historical) theories (or research programmes) and can be criticised by criticising the rational historical reconstructions to which they lead' (p. 21, italicised in the original). For a former student of Sir Karl Popper's the natural way to interpret 'criticise' is 'falsify', and that interpretation yields the viewpoint that Lakatos first explores. A methodology or 'rationality theory . . . is to be rejected', he suggests, 'if it is inconsistent with an accepted "basic value judgment" of the scientific élite' or if it fails to 'reconstruct the acknowledgedly best gambits as "scientific" ' (pp. 22 f., where the first sentence appears in italics). Knowledge of 'basic value judgments' or 'the acknowledgedly best gambits' of scientists is, of course, to be had from history, which thus becomes a vehicle for the critique of methodological theories.

In this first form, Lakatos quickly emphasises, his 'basic idea' will not quite do. Using well-known historical examples he systematically falsifies all four of the methodologies discussed in Part 1, his own included. Like scientific theories, he points out, methodologies can always be confronted with factual anomalies of the sort he has illustrated: 'criticise' cannot therefore be read in its most standard Popperian sense. But it can, Lakatos continues, be read in the more liberal sense developed for his own methodology of research programmes. Though methodologies may not usefully be falsified individually, they may be evaluated in pairs, the palm being awarded to the one that fits the larger range of historical facts. 'We then reject a rationality theory only for a better one', Lakatos writes, 'for one which, in this "quasi-empirical" sense, represents a ***progressive shift*** in the sequence of research programmes of rational reconstructions. . . . In the light of better rational reconstructions of science one can always reconstruct more of actual great science as rational' (pp. 30 f., italics in original). From this perspective, Popper's theory 'clearly represents progress over its justificationist predecessors, and in particular, over inductivism' (p. 30). The methodology of research programmes may, in turn, offer 'yet a further step forward' (p. 31). That, at least, is the position taken in Lakatos's essay, and the historical studies which follow are in part designed to evaluate it. Generally speaking, they aim to show that much scientific behaviour, previously seen as anomalous or irrational, demanding of external historical explanation, appears appropriate and rational when viewed in terms of Lakatos's methodology of research programmes.

As a motive for doing history that one seems to me a likely invitation to disaster. Dangers arise partly from the nature of historical research, partly from the way Lakatos suggests doing it. In both history and the sciences the selection and interpretation of sources of data are influenced by prior expectations, but in neither field does 'to influence' imply 'to determine' or 'to dictate'. Data can, and must be permitted to, react back on expectations, make trouble for them, play a role in their transformation. That mechanism is essential to the evolution of ideas in both fields, which are to that extent methodologically similar.

The mechanism operates differently in science and history, however, and the result is a decisive divergence in their optimal research strategies. In the sciences expectations are ordinarily quite precise. Though practitioners often differ about the significance of some particular anomaly, the existence of a departure from expectation is readily recognised and generally acknowledged. Since contributions to science often result from following up an anomaly or from demonstrating that none exists, the scientist who works without a firm grasp of the expectations determined by relevant theory is unlikely to succeed. In history, on the other hand, expectations are far less precise, and there is correspondingly less agreement than in science about whether expectations 'fit the facts' and about the sorts of data relevant to their evaluation. If prior expectations are not merely unreasonable, the historian committed to them can usually make a case without consciously forcing the data. Accordingly, the historian is usually well-advised to set expectations aside before beginning research. If science and method, for example, are the subjects, then both should be learned from the people under study not from later scientific and philosophical texts. That advice is, of course, a council of perfection: no one can entirely set aside thought patterns induced by prior experience and training; such patterns do influence research, which in any case could scarcely begin without them. But it is nonetheless essential that the attempt to unlearn them be made. The historian's problem is not simply that the facts do not speak for themselves but that, unlike the scientist's data, they speak exceedingly softly. Quiet is required if they are to be heard at all. That is a principal reason why I have myself resisted attempts to amalgamate history and philosophy of science though simultaneously urging increased interaction between the two.[1] History done for the sake of philosophy is often scarcely history at all.

Lakatos's view of history makes these intrinsic difficulties even more severe. For him, behaviour that does not fit methodological expectations belongs to, or is to be explained by, external history. Since the latter is 'irrelevant for the understanding of science', it may be ignored. Sometimes Lakatos even suggests that internal history be rewritten so that potential anomaly disappears: 'Internal history is not just a ***selection*** of methodologically interpreted facts', he writes, 'it may be, on occasions, their radically improved version. . . . One way to indicate discrepancies between history and its rational reconstruction is to relate the internal history ***in the text***, and indicate ***in the footnotes*** how actual history "misbehaved" in the light of its rational reconstruction' (p. 18, italics in original).[2] Following instructions like these, historians will often fail even to notice the 'misbehaviour' of actual history. If they do notice it, furthermore, they have no obligation to pause and consider its possible bearing on their narrative. The fact that it is 'misbehaviour' renders it a subject for external history and thus by definition irrelevant. Having saddled himself with the internal/external distinction Lakatos cannot allow anomalies even the limited heuristic functions in history that he grants them in science. As a result he deprives the historian of an essential, if also problematic, research tool.[1]

---

[1] *Cf.*, 'The relation between the History and the Philosophy of Science', in my [1977], pp. 3–20.

[2] Both in the paper under discussion and elsewhere, Lakatos illustrates this practice. My response discusses both the practice and the illustrations in more detail than is appropriate here.

These difficulties with Lakatos's analysis of history also infect his proposal for choosing between methodologies by comparing the corresponding internal historical narratives. Discussing such comparisons, he presupposes the existence of what he calls 'actual history', presumably the pool of all historical 'facts'. The better methodology, then, is the one that classifies more of the pool as internal rather than external. But 'actual history' of the sort Lakatos requires is a myth. The pool of all data potentially relevant to history cannot be either identified or scanned. More to the point, the data in most parts of the pool are not, until after much interpretation, the facts which appear in historical narratives. When Lakatos provides an historical case to illustrate the comparative merits of the methodology of research programmes, he is not selecting the elements of his internalist narrative from 'actual history' but creating them from often distant data or else choosing from the similar creations of earlier historians. Under those circumstances, it is not surprising that the story he tells makes essential use of elements that other methodologies would relegate to external history. It is by no means clear, however, that proponents of those methodologies would accept the elements of his narrative as simply factual, and it is upon that agreement that his demonstration depends. History is interpretative throughout.

Nevertheless, Lakatos's central intuition seems to me sound. The difficulties of his position derive from the apparatus in which that intuition is embedded. Very often, when one version of a historical development replaces another, the second makes essential use of evidence some of which was recognised as fact (and then sometimes seen as awkward) by proponents of the earlier position. Newton's alchemical writings, for example, were for years dismissed with remarks to the effect that, for all his brilliance in mathematics and mechanics, Newton could not in chemistry escape the climate of his times. Today those same alchemical writings are widely seen to reveal essential characteristics of the creator of the *Principia*, a man whose special gifts are now discoverable in a far larger range of the texts that he wrote. Or again, much scholarship has aimed for the last thirty years to provide an account of Galileo that would reconcile the apparent conflicts in his writing with respect to the role played by experiment in his research. Consensus has not yet been achieved, but Galileo is no longer either the father of experimental science or a man whose achievements can be understood as requiring only the reinterpretation of familiar observations. Though many disagreements continue, the Galileo portrayed by historians today fits a far larger fraction of the available data than was the case a generation ago. Lakatos is, I think, clearly right to suggest that improved historical narratives are often the ones that give a central role to a larger body of evidence, usually eliminating recognised or potential anomalies (his candidates for externalist explanation) in the process.

---

[1] On pp. 161–75 John Worrall outlines an interesting revision of Lakatos' prescription for the interaction between history and philosophy of science. His account requires that successful external explanation of anomalies play an essential role in the appraisal of a methodology, a change that seems to me both more drastic than Worrall supposes and also a considerable advance. His version of the role of history entirely escapes the immediately preceding criticism. But there is something about his position that Worrall is not seeing. Once external history becomes relevant to the appraisal of methodologies, the external/internal distinction need no longer be introduced prior to historical research. One can instead, as the following paragraphs suggest, simply do history.

That point can be made, however, without recourse either to a concept of 'actual history' or, more significantly, to an internal/external distinction governed by prior standards of rationality. For the historian, actual history is simply history that has actually been written or a selected subset thereof. One way of improving on it is to improve its fit to the range of facts already made accessible by interpretation. Others involve reinterpreting the existing data base or else extending it. All of these processes result in narratives that aim to say what occurred and to make it plausible, aims that require no prior decisions about what part of what occurred was rational, what not. Once the historian has provided such a narrative, the philosopher (sometimes the same person, but wearing a different hat) may examine it, asking about its significance for current doctrines concerning scientific method. If history is to have a chance of influencing those doctrines, however, such questions should be withheld, insofar as possible, until the task of the historian is complete.

Approaching *Method and Appraisal* with those predispositions, I did not anticipate much pleasure or instruction from the five historical essays that form its core. But experience has falsified my expectations. Three of the chapters, those by Peter Clark, John Worrall and Elie Zahar are valuable original contributions to the literature on their subjects. Alan Musgrave's essay on the Chemical Revolution, though it adds little to currently accepted doctrine, is nevertheless a thoroughly useful account, well organized, balanced, and judicious. Only Martin Frické's chapter, 'The rejection of Avogadro's hypotheses', seems to me to do less than justice to the sources, in this case mostly secondary sources, from which it derives. As a group, then, these case studies in the methodology of research programmes are of presumptive significance and interest to both philosophers and historians of science. The latter may be annoyed by the extent to which philosophical motives (and sometimes philosophical shibboleths) overwhelm historical narrative, but they should not ignore the substantive findings of at least the three principal essays listed above.

By virtue of their shared methodological premises, the five case studies are closely parallel in structure. Each is concerned with one or more research programmes, such a programme being understood as a coherent research tradition extended over time. Each characterises a research programme explicitly or implicitly in terms of its hard core—the unquestioned assumptions which endure through theory change—as well as in terms of its heuristic—the set of strategic directives which guide the selection of research problems and of techniques for solving them. And each, finally, is concerned to show that contemporary evaluations of competing research programmes depend upon whether the programmes are progressing (in a '*progressive problemshift*') or stagnating (in a '*degenerating problemshift*'). A research programme is in the first of these states 'as long as its theoretical growth anticipates its empirical growth, that is, as long as it keeps predicting novel facts with some success'. It is in the second or degenerating state 'if its theoretical growth lags behind its empirical growth, that is, as long as it gives only *post-hoc* explanations either of chance discovery or of facts anticipated by, and discovered in, a rival programme' (p. 11). These criteria of appraisal and ultimately of choice (interestingly criticised and elaborated by

Zahar on pp. 216–19) are central characteristics of the methodology of research programmes, for it is they which replace such standard evaluative criteria as confirmation and falsification. When they are used as guides to history, behaviour that had previously seemed irrational, in need of external historical explanation, can be seen instead as part of internal history, a fully rational response to the situation confronting scientists of the day. That, at least, is the fundamental claim made in all five of these case studies.

Peter Clark's chapter, 'Atomism *vs.* thermodynamics', provides a particularly useful illustration, especially because the ground it covers has been relatively little studied and Clark has had to work mainly from primary sources. At the start he reminds readers that opposition to kinetic theory in the last decades of the nineteenth century has usually been attributed to the spread of positivist philosophy, an external cause. He will argue, on the other hand, that in the quarter century after about 1880 the kinetic theory was degenerating, while its rival—phenomenological thermodynamics—was progressing. To make his case, Clark first characterises the atomic-kinetic research programme. Its hard core 'consisted simply of the proposition that: the behaviour and nature of substances is the *aggregate* of an enormously large number of very small and constantly moving elementary individuals subject to the laws of mechanics' (p. 45, italics in the original). Its heuristic consisted of four directives, two concerning the introduction of models and the gradual elimination of simplifications intended to facilitate computation, two others concerning the use of statistical distributions. Clark traces the programme thus described through the succession of theories due to Joule and Krönig, Clausius, Maxwell and Boltzmann. In the process he shows the endurance of the core, illustrates how the heuristic functioned, and demonstrates that, until around 1880, the succession of atomic-kinetic theories constituted a series of progressive problemshifts. Their adoption provided unexpected explanations of previously known effects, and it occasionally also permitted the prediction of previously unexpected regularities. During the following quarter century, Clark continues, that progressive pattern ceased. Novel problem solutions were no longer forthcoming, and in their absence two previously recognised anomales—the specific heat of gases and the incompatibility of statistical theory with the second law of thermodynamics—became more and more acute.

Clark turns next to the development of thermodynamics, tracing the subject from the Carnot memoir in the 1820s, through the work of Kelvin and Clausius in the 1850s and 1860s, and on to the work of Gibbs, Planck, Ostwald, Helmholtz, and others in the last three decades of the century. The hard core of the programme was, throughout, the existence of 'a definite relation between a quantity of heat and the quantity of work which in any way could be produced by it' (p. 63, italicised in the original). The heuristic emphasised the invention of idealised cyclic and reversible processes from which, together with a constitutive equation, functional relationships between empirical parameters could be derived. As in the case of kinetic theory, Clark is able to exhibit the thermodynamic programme as progressive, but in this case the progressive stage continued through the nineteenth century and into the twentieth, especially by virtue of the rapid development in and after the 1870s of chemical thermodynamics. It is these last developments, Clark suggests, not a renascent positivism, that led to the neglect of kinetic theory in the late nineteenth century. Rational judgment of technical promise, not the influence of an external philosophy, can be seen to account for the appraisals scientists made of the two competing programmes.

The other four papers differ from Clark's primarily in subject matter, scarcely at all in their goals. John Worrall's chapter, for example, is 'Thomas Young and the "refutation" of Newtonian optics'. It opens by rehearsing the three usual external explanations for the neglect of Young's wave theory—the authority of Newton, obscure presentation, and the vitriolic critiques published in the *Edinburgh Review*—and it then argues persuasively that none of them is plausible. As a substitute, Worrall presents a penetrating and often original analysis of Young's theory of interference and of the experiments presented in its defence. Neither the theory nor the experiments were quite so new as they have been taken to be, he points out, and neither supplied answers to the three primary objections that had been raised against waves-theories in the past. The few experiments which were new could, at least after the event, be explained by the corpuscular theory. In any case, Young's own explanations of many of them were seriously defective, and he was forced to adjust his stand again and again. In short, Worrall concludes, the wave optics programme never did progress in Young's hands, a sufficient reason and an internal one for it to have been ignored. The progressive problemshift which led quickly to the wave theory's triumph began only with the work of Fresnel. Martin Frické's explanation of the neglect of Avogadro's hypotheses takes a similar form, though its argument is neither so original nor so cogent. The outlines of Alan Musgrave's very satisfactory answer to the question of his title, 'Why did oxygen supplant phlogiston?', will be easily imagined by those who know the main recent literature of the Chemical Revolution.

Elie Zahar's 'Why did Einstein's Programme supersede Lorentz's?[1] requires special comment, for, though its aim is the same as that of its predecessors, its particular interest is due to the tools it develops on the way to that goal. Most previous accounts of the superiority of Einstein's special relativity theory to Lorentz's theory of corresponding states have emphasised that the latter was produced by *ad hoc* adjustments of classical theory simply to meet the results of the Michelson–Morley experiments. Against this standard viewpoint, Zahar argues with much success that Lorentz's successive theories were in no sense *ad hoc* and that their development until 1905 constituted a progressive problemshift. More important than that conclusion are two refinements that Zahar's arguments require: first in the notoriously vague notion of *ad-hoc*ness and, second, in the concept, vital to the methodology of research programmes, of what it is to be a novel fact. A second especially interesting aspect of Zahar's chapter depends upon the conclusion already reached. As extended by Poincaré in 1905 and 1906, Lorentz's theory was fully equivalent to Einstein's in its observational consequences. If, in addition, Lorentz's theory cannot properly be described as *ad hoc*, why did scientists like Planck and Minkowski rapidly abandon Lorentz's classical approach in favour of Einstein's? Zahar argues that the answer is the extra power of Einstein's heuristic, a conclusion the interest and strength of which depends upon the concrete detail he utilises in exhibiting heuristic power and in thus transforming a not-unusual intuition into an analytic demonstration. Zahar's paper then concludes by sketching a last stage in the relationship between Lorentz's theories and Einstein's. Only with the general theory of relativity, he points out, did Einstein's programme predict observations not also derivable from Lorentz's, for example the precession of Mercury's perihelion. And only after that occurred, does Zahar feel that Einstein's programme had genuinely superseded Lorentz's.

---

[1] Previously printed in The *British Journal for the Philosophy of Science*, **24** (1973), pp. 95–123, 223–62.

Given all that has previously been said about the likely consequences of doing history for philosophical purposes, how can these papers be so successful? One reason, clearly, is that their authors do not exploit the license that Lakatos's theory of history provides, but conform to more usual standards of historical responsibility. The occasional footnotes which aim to qualify the main text of these papers do not 'indicate . . . how actual history "misbehaved" in the light of its rational reconstruction' but instead point out that a fuller or less simple version of events would not change any points of current interest (*e.g.*, John Worrall on p. 130, n. 98, or compare Lakatos's remarks on p. 18 with those of Zahar on p. 219). But something more fundamental than respect for the historical record also helps to make these achievements possible. Reading these case studies, I was repeatedly struck by the fact that, though several provide novel historical information, the stories they tell are of a generally standard sort, unlikely to surprise anyone who has followed the better professional literature written during the last forty years about the development of scientific ideas. As history, that is, these case studies are relatively old fashioned, belonging to a tradition far older than the methodology of research programmes. For two decades, furthermore, that pre-existing historiographic tradition has provided the shared background for a number of historically oriented redirections of philosophy of science. Some of them have a good deal in common with the methodology of research programmes, and all of them should, I think, be seen as more nearly responses to than causes of a new way of doing history. Lakatos's expectations to the contrary, the case studies in *Method and Appraisal* do not illustrate the fructifying influence of the methodology of research programmes on historical practice.

It would, however, be strange if these studies had entirely escaped the influence of their authors' *parti pris*, and one recurrent aspect of the volume suggests that they have not done so. Except for the chapters by Musgrave and Frické, both based on secondary sources that consider the history of acceptance, these papers seldom present or even seek evidence of what actually attracted scientists to or repelled them from the various research programmes under study. Historically meticulous when dealing with the nature and development of scientific ideas, they substitute philosophical analysis for historical investigation when discussing why and how those ideas appealed. If analysis discloses a philosophically relevant difference between research programmes, then that difference is assumed to have played a role in programme choice.

One result is at least occasional errors and implausibilities. Peter Clark, for example, emphasises the irreconcilability of the kinetic theory both with the measured values of the specific heats of gases and with the phenomenological second law of thermodynamics. The first of these conflicts was a widespread source of concern by at least the 1890s, but the second appears to have bothered only Planck, and the evidence about even his attitude is somewhat equivocal.[1] Or again, Elie Zahar emphasises the equivalence of the Lorentz–Poincaré theory to special relativity but fails to ask how widely the equivalence was recognised during the first decade after Einstein's theory was introduced. That failure may be part of the reason why he insists that Einstein's programme only truly superseded Lorentz's with the invention of the general theory (p. 262). I am, however, quite sure that by or very shortly after 1915 special relativity had been professionally accepted and put to use to an extent that general relativity had not achieved even thirty or forty years later when I was myself a student. Acceptance might have occurred in the way Zahar suggests; the analytic differences he discloses with such skill were present. But they need not have played the role he attributes to them, and there is some reason to doubt that they did so.

A second difficulty that might have responded to the examination of actual appraisals is potentially far more serious. Lakatos's methodology attributes a virtually exclusive role in the evaluation of research programmes to what James Conant once called 'fruitfulness', the ability to evoke new discoveries. That criterion doubtless does play a role in actual appraisals, and the reasons for its use are clear. But fruitfulness is not the only criterion of appraisal, and its importance is, I think, decisively overstated in this volume. A primary result of that overstatement is a somewhat distorted view of the process of transition from one research programme to its successor.

Two oddities make what is at issue hard to see. Both for Lakatos and for me, the transition between incompatible research programmes takes the place of what, in more traditional methodologies, was the confirmation, verification, or falsification of an individual theory. That central aspect of Lakatos's viewpoint is, however, surprisingly little represented in his theory of appraisal. Though actual appraisals are for Lakatos always comparative, the single criterion in terms of which such appraisals are made is applicable to individual research programmes taken alone. Choosing between two research programmes, that is, one first locates each individually on the fruitfulness scale, the progressive/degenerative continuum, and only then does one compare the two. What one compares, furthermore, is simply their individual scores. Substituting the choice between programmes for the acceptance or rejection of an individual programme has had only a secondary effect on methodology.

That Lakatos's methodology seems nevertheless to work quite well is at least partly due to a corresponding oddity in the selection of case studies. Only two of the five, those by Musgrave and Zahar, illustrate the problem of choosing *between* programmes. Young's programme, unlike Fresnel's, was never a genuine challenge to corpuscular optics. Avogadro's hypotheses were not a research programme at all but a proposed addition to a programme already in existence. Kinetic theory and thermodynamics are compatible and could be held together by anyone who accepted a probabilistic second law. In cases like these, appraisal need only determine whether an individual programme is worth pursuing, and fruitfulness alone provides a sufficient basis.

When research programmes must be compared, however, when the decision to accept one is simultaneously the decision to abandon another, the situation is different in two essential respects. Competing research programmes need not in principle display differences in fruitfulness. Whether or not they do, furthermore, at least one other criterion must be considered when choosing between them. A preferred programme is expected not only to produce new achievements but also to conserve the achievements of the programme it replaces. In that situation, the successes and failures of each programme provide the standards against which the other must be evaluated. Appraisal then becomes intrinsically relative, involving both programmes as well as nature from the start. Musgrave's paper is the only one in *Method and Appraisal* to give a sense of the dynamics of the resulting competition. Instead of locating each programme separately and as a whole on the fruitfulness scale, Musgrave follows now standard historical practice and traces the changing responses over time of the oxygen and phlogiston programmes to a series of distinct chemical problems: combustion and calcination, acidity, and the composition of water. (I would myself have compared also the responses of the two programmes to the many contemporary discoveries of apparently distinct gases and gaseous combinations.) The picture that results is not easily fitted to Lakatos's model, partly because of its multidimensionality and especially because so many of the new phenomena that oxygen ultimately triumphed by explaining were first observed by phlogistonists. One could argue that the oxygen theory was itself the crowning demonstration of the progressiveness of the phlogiston programme.

---

[1] *Cf.* my [1978], pp. 18–29.

Those sentences isolate the single aspect of the methodology of research programmes that I find fundamentally unsatisfactory. (It is also, of course, one of the two primary differences between Lakatos's methodological position and my own, the other being the place of incommensurability.) Historically, when a choice must be made between two research programmes, the more recent one is usually an attempt to resolve concrete difficulties encountered in the practice of its older rival. In the sciences, unlike the arts and probably also unlike mathematics, new approaches do not ordinarily emerge simply when old ones stagnate, cease to produce novelty. Instead, they await a time when the traditional approach is seen to have failed in resolving problems acknowledged to lie within its competence.[1]

The difference is between what Lakatos calls 'stagnation' and what I have called 'crisis', and what is at stake can be clarified by comparing Lakatos's methodological views with those of Sir Karl Popper. Whatever its difficulties, Sir Karl's emphasis on falsification captured an aspect of scientific life that the methodology of research programmes effectively ignores. A scientific theory embodies expectations about natural phenomena; those expectations can be disappointed; and, when they are, some scientists usually start to look for an alternate theory. In ignoring this function of severe anomaly, Lakatos's position has, I think, moved too far from Sir Karl's. But in a closely related respect it remains too close. Once a theory has been falsified (Sir Karl) or entered a degenerating phase (Lakatos), it may be displaced by another theory. According to both these viewpoints, however, the existence of a potential replacement theory appears pure coincidence. Neither falsificationism nor Lakatos' methodology suggests how closely successive theories or research programmes relate to each other: the successor is usually conceived in response to, and is often also shaped by, particular difficulties encountered in the development of its predecessor.

---

[1] The respects in which mathematics is more like the arts than the natural sciences may have special relevance to this discussion. Lakatos was trained in mathematics, I in physics, and those bits of biography may help explain some differences in our positions. In mathematics as in art, traditions which stagnate do die out. (*Cf.*, Fisher, [1966] and [1967]). In the natural sciences, on the other hand, theories are either replaced or else preserved, stagnant or not, for their utility in application. Or again, crisis is the usual prelude to theory change in science but not in mathematics or in art. One cannot imagine an equivalent in the natural sciences for the famous unsolved problems—*e.g.*, Fermat's Last Theorem—which for a century or more challenged successive generations of mathematicians. In the sciences a puzzle that persistently resists the best efforts of the best professionals is taken to reflect an inadequacy of accepted theory. Finally, though they seldom read them as they were read by contemporaries, mathematicians are far more likely than natural scientists to know and discuss old mathematical works. (For the equivalent in the arts, see T. S. Elliot's classic essay 'Tradition and the Individual Talent'.) For reasons and in ways that remain obscure to me, the sciences destroy their past far more thoroughly than do mathematics or the arts. (Compare my 'Comment on the Relation of Science and Art', in my [1977], pp. 340–51.)

In the long run, the main importance of this pattern may prove to be the clues it provides to the sorts of continuity compatible with fundamental theory change. But it has a presumptive bearing as well on the problem of appraisal. In the competition between theories or research programmes, a special role is often played by the past history of the older theory, by difficulties recognised as recalcitrant or as sources of continuing disagreement even before the new research programme was proposed. In the case of the phlogiston theory, these were the difficulties over weight gain and also the problems of interpreting the increasing complexity of gas reactions, both of which had also played a key role in calling the oxygen theory forth. In the case of Lorentz and Einstein, I should suppose that many of the early converts to special relativity were attracted just because Einstein's theory at one stroke did away with—rather than elaborately resolved—a central and long recognised difficulty of classical theory. In this whole area, the methodology of research programmes can still learn from history.

Having written about *Method and Appraisal* at such length, I may not end without mentioning 'On the critique of scientific reason', the brief essay with which Paul Feyerabend closes the volume. Of course the author of *Against Method* is against the method presented in this collection. His essay supplies many reasons for his opposition, some brilliant, many cogent, and all unrelievedly negative. The last characteristic underscores what one already knew: Feyerabend does not want to improve methodology or the theory of science, but to do away with it. That aim he is unlikely to achieve, but it gives him the sceptic's easy advantage in debate. Any gap in a particular methodological argument is for him a sign of the impossibility of an entire enterprise. Response in this place would be pointless.

THOMAS S. KUHN
Massachusetts Institute of Technology

REFERENCES

FISHER, C. S. [1966]: 'The Death of a Mathematical Theory', *Archive for History of Exact Sciences*, 3, pp. 137–59.

Reviews

HOLTON, GERALD [1978]: *The Scientific Imagination: Case Studies*, Cambridge University Press. Hardback: £15. Paperback: £5.95 Pp. xvi+382.

Two-thirds of the essays in this volume have little to do with the process of scientific discovery. There are reviews of books by Lewis Mumford, Frank Manuel and Ronald Clark, an interesting critique of so-called 'science indicators', an externalist study of Fermi's research team in Italy, a review of old research done on the personality profiles of scientists, and a note on the assumptions underlying the Project Physics Course for high school students.

Holton believes that once lay people learn that scientists are not positivistic robots and that imagination plays an important role in scientific discovery much of their fear and hatred of the dehumanising effects of technology will dissipate. I suspect this is rather like trying to convince a lover of natural foods that DDT is 'organic' because it is made from coal-tar products which originally came from living trees. But I wish him every success in arguing with the folks he calls 'Dionysians', those gloomy enemies of science who place a high epistemological and moral value on immediate, unanalysed, subjective experience.

My primary concern here will be with the essays described by the title. Holton believes that contemporary philosophers of science, influenced by Popper and Reichenbach, have neglected the problem of scientific discovery and the role of extra-empirical elements in the growth of science. He makes no reference to the vast philosophical literature on prior appraisals, scientific heuristics, regulative principles, metaphysical research programmes, logic of pursuit, *etc.*, which decisively refutes his sociological claim of neglect.

But let us turn to Holton's own positive proposals. Some philosophers, including Einstein (whom Holton frequently quotes), have argued that since there is no logical path leading from experimental data to theory, theories are completely 'free inventions of the human intellect' (Einstein's phrase). Not so, claims Holton. The scientific imagination is often constrained by *themata* which predispose the mind towards certain theories rather than others. Different scientists (Bohr and Einstein are excellent examples) may be guided by opposing themata and hence have quite different intuitions about which sorts of theories will prove successful.

Holton distinguishes three ways in which themata influence science (p. 9). First, there are thematic *concepts*, such as symmetry and simplicity, which place constraints on acceptable scientific hypotheses. Secondly, there are *methodological* thema (such as a preference for analysis as opposed to synthesis) which partially determine how a scientist does research. Last are thematic *propositions*, such as the principles of special relativity, which form the core or starting point of a theory.

Holton *claims* (p. 23) that there are 'overwhelming differences' between his themata and good old-fashioned metaphysics, paradigms or world views, but I was unconvinced.

To investigate the role of themata in scientific discovery Holton suggests the use of case studies, but not the speculative rational reconstructions advocated by Lakatos. Rather we should follow Sir Peter Medawar's advice to listen at the laboratory keyhole. I will comment on Holton's historical methodology later.

The one case he presents in detail is intriguing, not as it turns out because of the themata employed, but because of the philosophical questions it raises concerning experimental error. The controversy beginning in 1910 over the charge on the electron (which Holton describes as 'a battle of two worlds') involved Millikan (a staunch atomist) and Ehrenhaft (who eventually became an admirer of Mach). Roughly, the situation was this: Millikan's oil-drop experiment yielded a discrete value of about 4.6×10^{-10} esu for the charge of the electron. Ehrenhaft's alternative method (which measured the horizontal motion of metal particles) yielded a whole spectrum of values which ranged as low as 5×10^{-11} esu.

Since no one to this day has successfully explained the results of Ehrenhaft's experiments (which were continued up until the 1940s) and since Millikan's data were not as unambiguous as sometimes claimed, Holton argues that it was the atomistic *thema* which tipped the balance in Millikan's favour, especially during the early years.

I have no doubts that the drive towards atomism is a very strong one, but judging from the evidence which Holton himself provides, it seems unlikely to have been a decisive factor in this case.

First of all, Millikan himself was not an across-the-board atomist—he opposed the hypothesis that light was quantized (p. 72). Secondly, Ehrenhaft started out with what Holton describes as 'an atomistic preference . . . as clear and explicit as Millikan's' (p. 47). And at least at the beginning of the controversy he was searching for sub-electrons, electrical quanta which were much smaller than the electron. What appears to have changed Ehrenhaft's mind was *not* the sudden intrusion of a Machian anti-atomistic thema but rather the growing recalcitrance of his data to an atomistic interpretation. Millikan's results, on the other hand, while not perfect, nevertheless showed a fairly steady convergence towards the classical value for the charge on the electron as he improved the apparatus. In addition, independent theoretical and experimental evidence for the existence of the electron was accumulating.

In his recounting of this particular episode, I feel that Holton has overemphasised the importance of extra-empirical factors. In his story of Fermi's discovery of the absorption of slow neutrons, on the other hand, I suspect that he underplays such considerations. The reason for both errors, I believe, stems from his ear-to-the-keyhole historiographic methodology.

In 1934, Fermi made the important and surprising discovery that neutrons which have been drastically slowed down (through collisions with hydrogen atoms) are extremely effective in inducing radioactivity. Fermi himself describes his decision to introduce a slab of paraffin in front of the neutron source as a piece of pure serendipity. Holton quotes Fermi: 'It was just like that: with no advanced warning, no conscious, prior reasoning! I immediately took some odd pieces of paraffin . . .' (pp. 155–6).

Holton, who is faithfully listening at the keyhole, believes Fermi, eschews speculative reconstructions, and concludes that themata are of 'no great help' in understanding Fermi's discovery (p. 22). He then launches into a study of the interpersonal relationships within Fermi's team, the rise of Italian nationalism, *etc.* But along the way Holton drops hints which with a little reconstruction might very well help us understand how Fermi arrived at his discovery. For example, the team had previously noted that experiments done on a wooden table led to more neutron absorption than those conducted on marble tables (p. 190). And Fermi had already worked on the theory of collisions of very slow electrons with atoms. As Amaldi later remarked (the quote is in Holton): '. . . this is exactly the same theory that was used one and a half years later for slow neutrons against nuclei . . .' (pp. 158–9).

If Holton would allow historians to be one-tenth as imaginative in working with their data as he argues scientists are, I think we might arrive at a better understanding of Fermi's break-through and scientific discovery in general.

NORETTA KOERTGE
Indiana University

BUTTS, ROBERT E. and PITT, JOSEPH C. (*eds.*) [1978]: *New Perspectives on Galileo.* Dordrecht–Boston: D. Reidel (University of Western Onrtario Series in Philosophy of Science, Vol. 14). Hardback: $29.50. Paperback: $12.50. Pp. xvi + 262.

This volume collects in revised form some of the papers presented at a workshop on Galileo held in Blacksburg, Virginia during October 1975. The title is somewhat misleading, since much of the material is not particularly new and even a larger percentage fails to shed new light upon Galileo. This said, the qualification must be made that there is still a reasonable amount of useful material for interpreting Galileo's place in the history of science and at least one of the papers is of outstanding importance.

Each of the editors contribute a paper: Butts on 'Some Tactics in Galileo's Propaganda for the Mathematization of Scientific Experience' and Pitt on 'Galileo: Causation and the Use of Geometry'. I find neither of them very illuminating, rooted as they are more in the twentieth-century philosophy of science than in the natural philosophy of Galileo's time. Butts once more takes up the hoary problem of Galileo's Platonism (p. 60) without attempting to push the analysis beyond the banalities of the textbook accounts and taking no note of the recent work which has shed new light on the subject such as the studies of Galluzzi ([1973]) and Purnell ([1972]) who have raised the level of discussion significantly. Moreover, Butt's forthright admission that he will pay no attention to Galileo as a historical figure (p. 82, n. 1) dooms his attempt from the outset. Such an approach is less reliable than that advocated elsewhere by Machamer (p. 164) and leaves one in a foggy world where fact and fiction blend together into a nebulous gray. Pitt hopes to understand Galileo by consulting a single English translation of one work and a couple of writings by philosophers who are by no means specialists on Galileo. Given such an impoverished diet there is little wonder that his results are less than convincing. What he says of Galileo's notion of science (p. 188), for example, has few of the rich nuances evident in McMullin's account elsewhere in the same volume. The paper contains the germ of a bright idea which has not been allowed to come to fruition owing to a lack of proper fertilisation through direct contact with the *Opere* of Galileo.

A number of other papers are more promising. H. E. Le Grand on 'Galileo's Matter Theory', while not really breaking any new ground, is a useful exposition, though one would have liked to see something about the extensive recent work on the problem of the continuum in the Middle Ages, especially that by J. E. Murdoch ([1962], [1964], [1969]). W. R. Shea's 'Descartes as a Critic of Galileo' is a valuable summary of the French philosopher's view of his Italian contemporary. Peter Machamer's 'Galileo and the Causes' argues that 'Galileo . . . might best be looked at as trying to legitimize his mathematical pursuits in the eyes of his university colleagues and the intellectual world at large by showing how the results of that tradition could be stated in traditional causal terms (p. 164)'. In spite of the many misprints marring the paper, there is the germ of something quite promising here, which could eventually lead to a genuinely 'new perspective on Galileo'. To make the case more convincing, however, the author must relate Galileo's thoughts on cause to what his contemporaries had to say on the subject. If Galileo was trying to 'legitimize' his endeavour, we need to have some sort of clear statement as what might have counted as legitimisation by his contemporaries. Also, I think it would be useful, should the analysis be carried further, if some consideration were given to Galileo's early works (*e.g.* the *De motu*) to see whether his thinking about causality and its role changed significantly as he matured.

The remaining three papers are significantly longer than the rest and certainly must be considered the centrepieces of the volume.

Ernan McMullin's 'The Conception of Science in Galileo's Work' traces two very different conceptions of science through Galileo's writings. The first is the Greek tradition of demonstrative science common both to the Aristotelian and the mathematical traditions. The second is what McMullin terms the 'retroductive notion of science'. The former, encompassing a much stronger and rigorous conception of science, was found by Galileo to be of limited application, being of use only in mathematics or in the relatively restrictive areas of natural science where the physical premisses are intuitively evident. This made necessary a more inductive approach based upon observation, which had wider application in cosmology, astronomy, and other fields of science. What the second type of science gained in breadth of application it lost in rigour. Indeed, for Galileo the second type was not science in the strict sense at all and throughout his life the Tuscan held that science, in the full meaning of the word, must have the requirements of the deductive science of Aristotle and the mathematicians. McMullin is quite right to conclude: 'We can hail Galileo and Descartes as the pioneers in the development of a new conception of science that would ultimately replace the older demonstrative one. But if we do, we ought to recognize how reluctant Galileo would have been to accept this honor' (p. 252). All in all his piece is provocative and informative dealing with a range of Galileo's writings, though I believe his approach to be a little out of contact with the real and immediate intellectual context in which Galileo was involved. It is as though Galileo were dialoguing directly with Aristotle and no one else, whereas we know from his polemical remarks that the number of his opponents was by no means small. Still, McMullin's attempt is to be taken seriously and should be a starting point for further discussion.[1]

Readers of Winifred Wisan's earlier work [1974] will know what to expect in her, Galileo's Scientific Method: A Re-examination', an extremely valuable- and up-to-date discussion of the *status quaestionis* of this much debated issue. Re-echoing an earlier statement by Crombie, she notes that a source for practically all philosophical and scientific methods can be found in Galileo if we look hard enough—and how hard historians and philosophers have looked! She is out to put it all back into perspective by tracing the approach to science which evolved as Galileo matured. She rightly emphasises that his approach to natural knowledge became more empirical—and, indeed, more experimental—as he grew older. Besides displaying an excellent command of the relevant primary sources, she is well aware of the recent interpretative work on the subject such as the noteworthy articles of Jardine ([1976]) and Galluzzi ([1973]) (see her excellent bibliography, pp. 54–7), so that our understanding of the methodological issues involved are on a much firmer basis than a few years ago when Randall's overly-influential [1940] paper still held sway.[1] There is much to recommend Wisan's careful analysis as one of the best starting points for a study of Galileo's thought as a whole. Among the few reservations I have about the paper is that I can see no convincing evidence to relate Galileo to scepticism (pp. 23–4). If Galileo is dubious about reaching certitude in all things, so too was Aristotle (*e.g. Nicomachean Ethics* I, 3). Scepticism involves more than momentary doubt and I know of no passage in Galileo's extant work indicating a systematic expression of doubt.

The paper which genuinely provides a 'new perspective' is W. A. Wallace's 'Galileo Galilei and the *Doctores Parisienses*'. It is also the only paper in the volume to provide a substantial amount of new information which clarifies some aspect of Galileo's life and works. It primarily involves an interpretation of manuscript 46 of the Galileo Collection of the Biblioteca Nazionale of Florence, which was published nearly a century ago in Favaro's edition (vol. I, pp. 15–177). This manuscript has puzzled Galileo scholars for many years. Although it has generally been thought to consist in a group of youthful writings—Favaro published it under the title of *Iuvenilia*—the content is so traditionally Aristotelian that it has been difficult to interpret the gulf between it and not only the *Two New Sciences* (1638) but even the *De motu* (*c.* 1590). Favaro, without really considering the rather scruffy and unappealing manuscript with much care, dated it 1584 and argued that the content derived from the teaching of Francesco Buonamici, whose enormous *De motu* appeared a few years later.[2] Shortly afterwards Pierre Duhem saw in the youthful writings certain reflections of the fourteenth-century Parisian school of natural philosophy and took them to give explicit confirmation to his thesis of the continuity of (French) medieval natural philosophy and early modern science.[3] Favaro himself

---

[1] See also Wallace [1976].

[2] Galilei, *Le Opere* (*ed.*): A. Favaro [1890–1909], 1, pp. 9–13.

[3] See Duhem [1913], especially 'La tradition parisienne et Galilée, III, pp. 562–83.

([1918]) responded to Duhem by arguing that the notes were derivative from rather pedestrian scholastic sources and, though undoubtedly in Galileo's hand, were not a product of his thought. Matters stood thus until a few years ago when Fr Wallace and Dr A. C. Crombie began to investigate the matter more carefully. Working independently (Dr Crombie in collaboration with Dr A. Carugo), they uncovered a new and somewhat surprising facet to the personality of the young Galileo. Dr Crombie and Dr Carugo have thus far reported their finding only in very cryptic and guarded form.[1] Fr Wallace has, however, in the paper being discussed here and in several other papers, put forward a rather remarkable bit of new information regarding the *Iuvenilia*.[2] They have discovered the sources from which the *Iuvenilia* derive, and the whole manuscript *is* derivative. The sources, however, are not Pisan as previously thought, but come in some way (which is not yet entirely clear) from the *ambiente* of the Jesuit Collegio Romano. The key figure linking Galileo with that institution is Christopher Clavius with whom Galileo had direct contact for many years and who served as a link to the material derived from Benito Pereira and others of the Collegio, some of which is still in manuscript. Wallace has obviously studied the question with great care and, though there are still some loose ends and unanswered questions, there is little doubt that he is on the right track and his competent research has done much to clarify a rather obscure aspect of Galileo's life before he went to Padua in 1592. Incidentally, he convincingly redates the *Iuvenilia* to about 1589, when Galileo was 25 years of age. The sources of Galileo's work of that period, as evidenced in the *Iuvenilia* are not from a predominantly nominalist context, as some had previously supposed, but from a Jesuit–Thomistic one, as strange as this might sound. Wallace's hypothesis concerning the purpose of the notes, which seems eminently sensible to the reviewer, is that they were prepared in the hope that their compiler might be able to move from his mathematics lectureship (which he got in 1589) to a higher paying position teaching natural philosophy.

This paper revises a good deal of what we know of the young Galileo and puts us in a much better position to understand various aspects of his early life. What it does not do, however—and in my view this still remains the major problem of Galileo scholarship—is to explain adequately how he emerged from the position represented by the *Iuvenilia* to become 'the father of modern science' a few decades later. The problem of the development of his thought from the conventional young man uncovered by Wallace through the iconoclastic but inconclusive *De motu* to the originality of the *Two New Sciences* must still be faced. The results of several papers in the present volume (*e.g.* those of Wisan, McMullin and Machamer) suggest that a full-scale study of the evolution of his thought should be undertaken by someone young enough to complete it and well-prepared enough to deal with the complex range of materials.

After not progressing very much for some years Galilean studies have flourished significantly during the past decade or so. Some of the new progress is charted here. In addition to the new light shed on the early Galileo by Wallace's research, the other important recent development has been the clarification of Galileo's use of experimental method, especially through the work of S. Drake, R. Naylor, T. Settle and W. Wisan. Even if they each interpret the material somewhat differently and the precise meaning of the manuscript material is not always clear, it cannot be denied that they are on to something important. It is unfortunate that this important new direction of Galileo research —perhaps the most important of recent years—has not been adequately dealt with in *New Perspectives on Galileo*. For that reason this volume is not as definitive statement of the 'state of play' from the late 1970s as we might have hoped for. Nevertheless, there is a good deal of meat here and several of the articles will be widely read and debated in years to come.

---

[1] See their contributions, pp. 157–95 and 303–5 in Bonelli and Shea (*eds.*) [1975].

[2] See Wallace [1974], and see also Wallace [1976].

C. B. SCHMITT
The Warburg Institute, University of London

REFERENCES

DUHEM, P. [1913]: *Études sur Léonard de Vinci*. Paris.

FAVARO, A. [1918]: 'Galileo Galilei e i *Doctores Parisienses*'. *Rendiconti della R. Accademia dei Lincei*, **27**, pp. 3–14.

GALLUZZI, P. [1973]: 'Il "platonismo" del tardo Cinquecento e la filosofia di Galileo' in P. Zambelli (*ed.*) *Ricerche sulla cultura dell'Italia moderna*, pp. 37–79. Rome.

JARDINE, N. [1976]: 'Galileo's Road to Truth and the Demonstrative Regress', *Studies in the History and Philosophy of Science*, **7**, pp. 277–318.

MURDOCH, J. E. [1962]: *Rationes mathematice: un aspect du rapport des mathématiques et de la philosophie au moyen âge*. Paris.

MURDOCH, J. E. [1964]: 'Superimposition, Congruence and Continuity in the Middle Ages', in *Mélanges Alexandre Koyré*, pp. 416–41. Paris.

MURDOCH, J. E. [1969]: *Mathesis in Philosophiam scholasticam introducta*: The Rise and Development of Application of Mathematics in Fourteenth-Century Philosophy and Theology', in *Arts Libéraux et philosophie au moyen âge*, pp. 215–54. Montréal-Paris.

PURNELL, F. [1972]: 'Jacopo Mazzoni and Galileo', *Physis*, **14**, pp. 273–94.

RANDALL, J. H. [1940]: 'The Development of Scientific Method in the School of Padua', *Journal of the History of Ideas*, **1**, pp. 177–206.

RIGHINI BONELLI, M. L., and SHEA, W. R. (*eds.*) [1975]: *Reason, Experiment and Mysticism*. New York.

WALLACE, W. A. [1974]: 'Galileo and the Thomists' in A. Maurer (*ed.*): *St. Thomas Aquinas Commemorative Studies* 1274–1974, **2**, pp. 293–330. Toronto.

WALLACE, W. A. [1976]. 'Galileo and Reasoning *ex suppositione*', *Boston Studies in the Philosophy of Science*, **32**, pp. 79–104.

WISAN, W. L. [1974]: 'The New Science of Motion: A Study of Galileo's *De motu locali*', *Archive for the History of Exact Sciences*, **13**, pp. 103–306.

ULLMAN-MARGALIT, Edna [1977]: *The Emergence of Norms*. Oxford University Press. £8.00. Pp. xiv+206.

The scientist who attempts to construct theories predicting people's actions faces the difficulty that people's actions are often influenced by their predictions about *each other*'s actions. For example, a judge's actions relative to a prisoner, or a parent's actions relative to a child, will be based on predictions about the effect of such actions on the actions of the other. Meanwhile, of course, the child's actions may be based on predictions of their effect on the parent's actions, and so forth.

For the scientist, this gives rise to the question: Would people's actions be affected if they generated their predictions from my (supposedly correct) theory about people's actions? If the answer is yes, then the theory has to be accompanied by some such self-referential qualification as: this theory is true at least for people unaware of its truth. It is then natural to ask: what actions can be expected of people *aware* of the theory's truth?

Since application of a theory tends to disseminate it, this question is particularly important for an applied theory.

As soon as one talks in this way, one is clearly talking game theory. In fact, I would suggest that game theory is just the discipline that attempts to give general answers to these general questions.

Let me say why I think this describes game theory. Firstly, it is an accepted principle of game theory that one should always check a theory by asking: Would it be falsified if the players became aware of it? Secondly, one may change any social-science model into a game simply by asking this question concerning it. Take, for example, a Richardson (differential-equation) model of an arms race. To answer the above question would be to take certain parameters of the Richardson process and assume them to be controllable by the actors (nations). This would be to assign the "players" strategies. Players' preferences would have to be based on some further assumptions. The result would be a game based on the Richardson model.

Or consider a Phillips model showing an inverse relation between unemployment and inflation. To put the game-theoretic question to this model one would have to ask three questions. (*a*) *Who* becomes aware of the theory? The answer gives us our players. A possible answer: the unions and the government. (*b*) *What* (that they are aware of) can they do about it? The answer gives us the players' strategies. Here strikes, wage-freezes and tax-cuts might be components of players' strategies. Note that if only one player can do anything, we have the degenerate case of a one-person game. If none can do anything, our question has been answered. Our original theory is proof against players' awareness of it, since in any case they are impotent. (*c*) Given that players have strategies by which to affect outcomes, what are their *preferences* between outcomes? The answer enables us to use game theory to predict their actions, so confirming or disconfirming the original theory.

Game theory as so described should have great scope. It should complement the investigation of *unintended* social consequences that Popper ([1957]) has recommended by investigating the relation between actions and intentions. This relation is trivial in the one-person case. A lone player will intend as a result of his/her action a consequence that he/she most prefers. But in many-person theory this "rationalistic" (utility-maximising) assumption becomes inadequate, even misleading. This has been argued particularly by Rapoport ([1964]) and myself ([1971]). Hence the need for game theory proper.

At the same time, game theory should remain a structural approach. It should be parasitic on the other social sciences, merely analysing their content in a certain way.

To me a particular appeal of game theory is that, by asking its characteristic question, it rejects the "scientistic" ideal which sees the scientist as outside and above the world of his subjects, whom he is thereby free to manipulate and exploit.

It is true, as it should be, that those who have persistently asked the game-theoretic question have been game theorists. But the answers given by mainstream game theory have not, I think been satisfactory.

What has been wrong with these answers? (*a*) They have, I think, been narrowly dogmatic and exclusive in offering unique "solutions" to games. These solutions are invariably derived by arbitrarily *presupposing* that no psychological or social factors not found in the abstracted game model can affect real-world outcomes. Such is the content, to take two examples, of the "symmetry" axioms used to derive the unique Nash bargaining solution and the unique Shapley value. It is clearly objectionable, given the above-noted parasitic relationship between game theory and the other social sciences, wherein game theory uses their results to build its models. It is not for game theory arbitrarily to *presuppose* that much of what they supply will be irrelevant. (*b*) Game-theoretic analyses have been limited by the principle that players' preferences must be taken as fixed and given—whereas in real life we see that playing the game, or bargaining over it, can affect emotions, values, and preferences. (*c*) Game theorists have mostly used quantitative mathematics (employing real-valued utilities and probabilities), whereas in applications and other disciplines the problems often appear as non-quantitative. Non-quantitative mathematics, being more general, is thus more appropriate for game theory. (*d*) Instead of seeing game theory as the exploration of irrational and non-rational (as well as rational) behaviour, mainstream theorists have persistently tried to extend the prescriptive mode of one-person theory to game theory proper. This has resulted in attempts to ignore or evade the paradoxes of game theory. Take, for example, the famous game of prisoner's dilemma, in which if each player (a prisoner) maximises his utility by betraying the other each is worse off (in utiles) than if each *minimises* his utility by keeping mum. To evade this dilemma, the structure of which however pervades all forms of social life, game theory splits itself into "cooperative" and "non-cooperative" theory, providing two contradictory theories in each of which two different "solutions" to the dilemma are maintained.

Game theory must, I think, correct these faults. That is why I find Edna Ullmann-Margalit's *The Emergence of Norms* tremendously interesting and welcome.

Ullmann-Margalit is not a game theorist or a mathematician. She is a philosopher. But she has posed a version of the game-theoretic question. She has therefore used game models. On looking at well-known, mainstream game theory, however, she has found it necessary to reject it for just the reasons I have outlined. So she calls what she has done "not game theory". But it is game theory in a wider sense. Whatever, it is exciting, path-breaking work.

Ullmann-Margalit sets out to explain the emergence of some social norms. This she does by pointing to frequently-occurring paradoxical situations (modelled by paradoxical games) in which rationality (utility maximisation) is found to be an inadequate or unfortunate principle. Social norms, she argues, are promulgated to overcome such problems. They do so by changing players' preferences—it being understood that ethical and other norms, such as the social norms being considered, are factors in determining preferences. The induced change in preferences gets round the problem by changing the game model to one in which the paradox no longer occurs.

Her method works. Three basic games, each with variants and extensions, are examined as potential norm-generators. The first is prisoner's dilemma. Among prisoners, this clearly generates the norm against stool pigeons. When a many-person version of the same game is played in military situations (where the non-cooperative choice for each soldier is to try to save his own skin) it generates norms of military discipline and honour. Elsewhere, it generates the norms according to which one should pay taxes and vote.

The last case is particularly poignant. Few would vote, since a single vote is without effect and inconvenient, were it not for the social norm that rewards one for voting by a feeling of having done one's duty. Thus the generation of this PD (prisoner's dilemma) norm is essential for large-scale democracy.

Ullmann-Margalit proceeds next to expose the foolishness of much work done by ethical philosophers since Kant on the "generalisation argument". This asks: What would happen if everyone refused to do that? It seeks to show that everyone *should* "do that". In some cases it seems weighty—as when to "do that" is to vote. Elsewhere it seems to fail—as when to "do that" is to grow food. Clearly it would be almost as bad for everyone to grow food as for no one to do so.

Why does the argument sometimes work and sometimes not? Clearly, it works when we are all in a PD situation. Then it simply propounds a PD norm. When our situation is not quite a PD, the argument doesn't quite work, and so on.

Here I should begin to point out how Ullmann-Margalit's work could be improved. Norms that are "not quite" PD norms are clearly generated by situations that are "not quite" PD. Why not deal with these? To do so, Ullmann-Margalit should not have taken a single simple game, such as prisoner's dilemma, as a "paradigm" for a wide class of situations. Instead, she should have produced the mathematical characterisations of wide classes of games that would correctly model wide classes of situations. But this is for the future. Game-theoretically, Ullmann-Margalit is sloppy. But she is not a mathematician. She has done what she could, and it is an excellent rough draft.

I will now postpone further criticisms. The next games Ullmann-Margalit considers are "coordination games". Typical of these is the game of choosing which side of the road to drive on—left or right? My choice matters little, provided it is the same as everyone else's. Hence there is a basic indeterminacy; rationality is not enough. To solve this problem there emerges a norm (ethically neutral this time) that says: drive on the xth side. Similar norms emerge to regulate currency, language and other conventions.

Finally (and somewhat less successfully) Ullmann-Margalit considers "norms of partiality". These arise from two-person games with two equilibrium points. At one equilibrium point, one of the players is favoured; at the other, the other is. Game theorists call such games "marriage games" because of a story about a husband and wife who have to decide whether to go to the opera (preferred by the wife) or a football match (preferred by the husband). Ullmann-Margalit takes them as modelling the social conflict between the ruling class and the ruled. In this conflict the "status quo" is at the equilibrium point favouring the ruling class. For the workers to move unilaterally away from here brings about an outbreak of class war—*e.g.* (presumably) a general strike or a revolution. How does this situation generate norms? The ruling class, being dominant in society, promulgates norms that assert the legitimacy of its favoured position against the ever-present threat of the alternative equilibrium—that favouring the workers. These are the "norms of partiality". The obvious reference is to Marx and other conflict theorists such as Dahrendorf.

This analysis again reaches its goal, I think. But there are now serious defects in Ullmann-Margalit's method of getting there. (*a*) She unnecessarily introduces interpersonal comparisons of utility. As she points out, these have generally been avoided by game theorists because they were not needed. She thinks, however, that she needs them, but as far as I can see her argument in no way depends on them. (*b*) Similarly, she unnecessarily introduces a very un-game-theoretic distinction between players' "payoffs" and their "preferences". She does so by supposing players' preferences to be influenced not only by their own "payoffs", but by a comparison between their own and other players' "payoffs". This is very confused. Game-theoretic preferences are represented by payoffs and are assumed to reflect unselfish motives such as envy and generosity as much as selfish ones. By no means does game theory "rule them out", as she asserts. But again the point is that the distinction is by no means necessary for her argument, which would be much clearer without it. (*c*) Ullmann-Margalit's derivation of the need for a partiality norm is also confused, being based on an unsatisfactory game-theoretic concept of her own called "strategic" (as distinct from "game-theoretic") "stability". (*d*) In the end her partiality norm induces a change in players' preferences which seems not to go far enough. Let me apply what she says to a possible general strike called to force the institution of socialism. Ruling-class propaganda, she seems to say, will attempt to make the workers dislike the general strike (which would be called for strategic reasons, not because it is liked) even more than they do already. But surely it will go further. Surely it will attempt to turn them against socialism itself (the alternative equilibrium which they hope to bring about). Curiously, she does not seem to think it will.

I could list other faults in this book. Generally, her faults arise from having rejected—or failed to understand—work done using rigorous mathematics. (For example, she seems to have entirely missed the point of Arrow's work on social choice theory.) I prefer to see these faults as the obverse of the book's merits. Ullmann-Margalit rejects game theory because of the faults in it that I have mentioned. Engrossed in "solution concepts" accurate to several decimal points, many game theorists seem, in comparison with her, like a convoy of blind men. She has turned away from them and made discoveries. It is unfortunate that in doing so she has rejected certain disciplines and tools that would have made her work stronger. But it is understandable. And it is certainly possible that others will extend and strengthen her approach, if she does not do it herself.

NIGEL HOWARD

REFERENCES

Howard, N. [1971]: *Paradoxes of Rationality*. Cambridge, Mass.
Popper, K. R. [1957]: *The Poverty of Historicism*. London: Hutchinson.
Rapoport, A. [1964]: *Strategy and Conscience*. New York: Harper and Row.

Brit. J. Phil. Sci. 31 (1980), 204–211 *Printed in Great Britain*

Recent Publications

Allen, G. E. [1979]: *Thomas Hunt Morgan—the Man and His Science.* New Jersey: Princeton University Press. £17.60. Pp. 447.

Allen, G. [1979]: *Life Science in the 20th Century.* London: Cambridge University Press. Paperback: £2.95. Pp. 257.

Alther, E. [1979]: *Das Absolute als Zeit-Raum-Verhältnis und Vorgang.* Zurich: Kris-Verlag Zurich. S.Fr. 37.90. Pp. 84.

Armstrong, D. M. [1978]: *Universals and Scientific Realism, Vol. I: Nominalism and Realism.* Cambridge: Cambridge University Press. £6.75. Pp. 149.

Armstrong, D. M. [1978]: *Universals and Scientific Realism, Vol. II: A Theory of Universals.* Cambridge: Cambridge University Press. £7.50. Pp. 190.

Asquith, P. A. and Hacking, I. (*eds.*) [1978]: *PSA 1978, Vol. I.* Michigan: Philosophy of Science Association. Hardback: $9.00. Paperback: $6.00. Pp. 314.

Axelrod, C. D. [1979]: *Studies in Intellectual Breakthrough: Freud, Simmel, and Buber.* Massachusetts: University of Massachusetts Press. $10.00. Pp. 93.

Baldamus, W. [1976]: *The Structure of Sociological Inference.* London: Martin Robertson and Co. £7.45. Pp. 238.

Ballard, E. G. [1978]: *Man and Technology.* New Jersey: Humanities Press. $12.50. Pp. 251.

Bambrough, R. [1979]: *Moral Scepticism and Moral Knowledge.* Henley: Routledge and Kegan Paul. £5.50. Pp. 166.

Barndorff-nielson, O. [1978]: *Information and Exponential Families in Statistical Theory.* Chichester: John Wiley and Sons Ltd. £13.50. Pp. 238.

Barwise, J. (*ed.*) [1978]: *Handbook of Mathematical Logic.* Amsterdam: North Holland Publishing Co. $75.00. Pp. 1165.

Bateson, G. [1979]: *Mind and Nature—A Necessary Unity.* London: Wildwood House Ltd. £7.50. Pp. 238.

Baumel, H. [1978]: *Biology: Its Historical Development.* New York: Philosophical Library Inc. $6.00. Pp. 101.

Bayon, D. and Gasparini, P. [1979]: *The Changing Shape of Latin American Architecture.* Chichester: John Wiley and Sons Ltd. £14.30. Pp. 254.

Berghel, H., Hubner, A. and Kohler, E. (*eds.*) [1979]: *Wittgenstein, The Vienna Circle and Critical Rationalism.* Dordrecht: Reidel Publishing Co. D.Fl. 95. Pp. 550.

Bernsen, N. O. [1979]: *Knowledge—A Treatise on Our Cognitive Situation.* Odense: Odense University Press. Paperback: Danish Kroner 120. Pp. 686.

Bhaskar, R. [1979]: *The Possibility of Naturalism.* Hassocks: Harvester Press. Hardback: £11.50. Paperback: £4.50. Pp. 228.

Bieri, P., Horstmann, R.-P., and Kruger, L. [1979]: *Transcendental Arguments and Science.* Dordrecht: D. Reidel Publishing Co. Pp. 314.

Blom, J. [1979]: *Descartes: His Moral Psychology and Philosophy,* Hassocks: Harvester Press. £11.50. Pp. 288.

Boden, M. [1978]: *Purposive Explanation in Psychology.* Hassocks: Harvester Press. Hardback: £10.50. Paperback: £4.95. Pp. 408.

Bogdan, R. (*ed.*) [1979]: *Patrick Suppes.* Dordrecht: D. Reidel Publishing Co. D.Fl. 22.50. Pp. 226.

Bowen, K. A. [1978]: *Model Theory for Modal Logic.* Dordrecht: D. Reidel Publishing Co. D.Fl. 50. Pp. 128.

Boyer, D. L., Grim, P. and Sanders, J. T. (*eds.*) [1979]: *The Philosopher's Annual.* Oxford: Basil Blackwell. £10.00. Pp. 223.

Brittan, G. G. [1978]: *Kant's Theory of Science.* Princeton: Princeton University Press. £9.10. Pp. 215.

Bronowski, J. [1978]: *Magic, Science and Civilisation.* New York: Columbia University Press. £8.70. Pp. 88.

Bronowski, J. [1979]: *The Commonsense of Science.* London: Heinemann Educational Books. £1.50. Pp. 159.

Bugliarello, G. and Doner, D. B. [1979]: *The History and Philosophy Technology.* London: University Press. £12.25. Pp. 384.

Bunge, M. [1979]: *Treatise on Basic Philosophy Vol. 4.* Dordrecht: D. Reidel Publishing Co. D.Fl. 90. Pp. 314.

Bursen, H. A. [1978]: *Dismantling the Memory Machine.* Dordrecht: D. Reidel Publishing Co. D.Fl. 40. Pp. 153.

Busch, C. [1978]: *Starnberger Studien 1: Die gesellschaftliche Orientierung des wissenschaftlichen Fortschritts.* Frankfurt: Suhrkamp Verlag. Pp. 417.

Cappelletti, A. [1979]: *Ensayos sobre los Atomistas Griegos.* Caracas: Sociedad Venezuela de Ciencias Humanas. Pp. 161.

Caroe, G. M. [1978]: *William Henry Bragg 1862–1942—Man and Scientist.* London: Cambridge University Press. £8.95. Pp. 212.

Calabresi, G. and Bobbitt, P. [1978]: *Tragic Choices—The Conflicts Society Confronts in the Allocation of Tragically Scarce Resources.* New York: W. W. Norton and Co. Inc. Pp. 252.

Calder, N. [1979]: *Einstein's Universe.* London: BBC Publications. £6.25. Pp. 178.

Cargile, J. [1979]: *Paradoxes—A Study in Form and Prediction.* Cambridge: Cambridge University Press. £12.50. Pp. 308.

Caws, P. [1979]: *Sartre.* Henley: Routledge and Kegan Paul. £8.95. Pp. 210.

Churchland, P. M. [1979]: *Scientific Realism and the Plasticity of the Mind.* Cambridge: Cambridge University Press. £8.50. Pp. 157.

Clegg, S. [1979]: *The Theory of Power and Organisation.* Henley: Routledge and Kegan Paul. £7.50. Pp. 176.

Cohen, I. B. [1978]: *Introduction to Newton's 'Principia'.* Cambridge: Cambridge University Press. £7.95. Pp. 380.

Cohen, I. B. (*ed.*) [1978]: *Isaac Newton's Papers and Letters on Natural Philosophy.* London: Harvard University Press. £18.55. Pp. 539.

Cohen, R. S. (*ed.*) [1978]: *Eino Kaila: Reality and Experience: Four Philosophical Essays.* Dordrecht: D. Reidel Publishing Co. Hardback: D.Fl. 110. Paperback: D.Fl. 55. Pp. 322.

Cohen, R. S. and Stachel, J. J. (*eds.*) [1977]: *Selected Papers of Leon Rosenfeld.* Dordrecht: D. Reidel Publishing Co. Hardback: D.Fl. 185. Paperback: D.Fl. 70. Pp. 922.

Cooper, W. S. [1978]: *Foundations of Logico-Linguistics—A Unified Theory of Information, Language and Logic.* Dordrecht: D. Reidel Publishing Co. Hardback: $26.00. Paperback: $12.50. Pp. 244.

Coulter, J. [1979]: *The Social Construction of Mind.* London: Macmillan Press. £10. Pp. 190.

Cremaschi, S. [1979]: *L'Automa Spirituale.* Milano: Vita e Pensiero. L. 8,000. Pp. 174.

Crescini, A. [1978]: *Il Senso Della Ricerca Scientifica.* Rome: Edizioni Dell'Ateneo and Bizzarri. L. 12,500. Pp. 420.

Crosland, M. [1978]: *Gay-Lussac.* Cambridge: Cambridge University Press. £15.00. Pp. 333.

Curley, E. M. [1978]: *Descartes Against the Skeptics.* Oxford: Basil Blackwell. £9.50. Pp. 242.

Dauben, J. W. [1979]: *Georg Cantor: His Mathematics and Philosophy of the Infinite.* London: Harvard University Press. £19.95. Pp. 404.

Davies, P. C. W. [1979]: *The Forces of Nature.* Cambridge: Cambridge University Press. Hardback: £12.00. Paperback: £4.25. Pp. 246.

Debus, A. [1979]: *Man and Nature in the Renaissance.* Cambridge: Cambridge University Press. Hardback: £7.95. Paperback: £2.50. Pp. 158.

Del Valle, A. B. F. (*ed.*) [1977]: *Humanitas—Anuario del Centro de Estudios Humanisticos.* Mexico: Universidad Autonoma de Nuevo Leon. Pp. 645.

De Martino, D. and Bianca, M. (*eds.*) [1979]: La Mente e la Macchina. Brescia: Editrice la Scuola. L. 3000. Pp. 145.

Derrida, J. [1979]: *Writings and Difference.* Oxon: Routledge and Kegan Paul. £12.50. Pp. 342.

Dumitriu, A. [1977]: *History of Logic. Vols. I–IV.* Tunbridge Wells: Abacus Press. £42.50 complete set or £11 per volume. Pp. 342, pp. 265, pp. 393 and pp. 275.

Durbin, P. T. (*ed.*) [1978]: *Research in Philosophy and Technology, Vol. 1.* Connecticut: JAI Press Inc. £17.95. Pp. 390.

Ehrenberg, W. [1977]: *Dice of the Gods—Causality, Necessity and Chance.* London: Birkbeck College. £2.00. Pp. 109.

Ellis, B. [1979]: *Rational Belief Systems.* Oxford: Basil Blackwell. £6.95. Pp. 118.

Elster, J. [1978]: *Logic and Society—Contradictions and Possible Worlds.* Chichester: John Wiley and Sons Ltd. £10.95. Pp. 235.

Elster, J. [1979]: *Ulysses and the Sirens.* Cambridge: Cambridge University Press. £9.75. Pp. 193.

Erwin, E. [1979]: *Behaviour Therapy.* Cambridge: Cambridge University Press. Hardback: £12.50. Paperback: £4.25. Pp. 250.

Feyerabend, P. K. [1978]: *Der Wissenschaftstheoretische Realismus und die Autorität der Wissenschaften.* Wiesbaden: Friedr. Vieweg and Sohn. Paperback: DM 34. Pp. 367.

Feyerabend, P. [1978]: *Science in a Free Society.* London: New Left Books. £7.50. Pp. 221.

Fraser, J. T. [1978]: *Time as Conflict.* Basel: Birkhäuser Verlag. S.Fr. 48. Pp. 356.

Frege, G. [1979]: *Posthumus Writings.* Oxford: Basil Blackwell. £15. Pp. 288.

French, A. P. (*ed.*) [1979]: *Einstein: A Centenary Volume.* London: Heinemann Educational Books, Ltd. £10.00. Pp. 332.

Gaston, J. [1978]: *The Reward System in British and American Science.* Chichester: John Wiley and Sons Ltd. £13.40. Pp. 200.

Geison, G. [1978]: *Michael Foster and Cambridge School of Physiology: The Scientific Enterprise in Late Victorian Society.* New Jersey: Princeton University Press. £18.40. Pp. 401.

Gellner, E. [1979]: *Words and Things—An Examination of and an Attack on Linguistic Philosophy.* Henley: Routledge and Kegan Paul. Hardback: £7.95. Paperback: £4.50. Pp. 287.

George, F. H. [1979]: *Philosophical Foundations of Cybernetics.* Tunbridge Wells: Abacus Press. £13.50. Pp. 155.

Goldmann, L. [1979]: *Lukacs and Heidegger—Towards a New Philosophy.* Oxon: Routledge and Kegan Paul. Paperback: £2.50. Pp. 112.

Goldstein, M. and Goldstein, I. [1978]: *How We Know.* New York: $14.95. Pp. 357.

Goni Arregui, F. [1979]: *La Cara Oculta del Mundo Fisico.* Madrid: Editorial Gredos. Pp. 238.

Grayson, H. W. [1978]: *The Theory of Relativity Revisited.* Ardmore: Dorrance and Co. Inc. $10.00. Pp. 261.

Greene, F. [1978]: *Peking.* London: Jonathan Cape. £6.95. Pp. 162.

Haack, S. [1978]: *Philosophy of Logics.* Cambridge: Cambridge University Press. Hardback: £13.50. Paperback: £3.95. Pp. 276.

Hall, A. R. and Boas Hall, M. (*eds.*) [1978]: *Unpublished Scientific Papers of Isaac Newton.* Cambridge: Cambridge University Press. Paperback: £7.95. Pp. 416.

Hamilton, A. G. [1978]: *Logic for Mathematicians.* Cambridge: Cambridge University Press. Hardback: £12.50. Paperback: £4.95. Pp. 224.

Harrison, A. [1978]: *Making and Thinking.* Hassocks: Harvester Press. £11.50. Pp. 207.

Hawking, S. and Israel, W. (*eds.*) [1979]: *General Relativity.* Cambridge: Cambridge University Press. £37.50. Pp. 919.

Heidegger, M. [1978]: *Basic Writings—Nine Key Essays: Introduction to Being and Time.* Henley: Routledge and Kegan Paul. Hardback: £7.50. Paperback: £4.50. Pp. 392.

Heilbron, J. L. [1979]: *Electricity in the 17th and 18th Centuries.* London: University of California Press. £24.00. Pp. 606.

Heitsch, W. [1978]: *Mathematik und Weltanschauung.* Berlin: Akademie-Verlag. DM. 19.50. Pp. 348.

Hicks, J. [1979]: *Causality in Economics.* Oxford: Basil Blackwell. £5.95. Pp. 124.

Hintikka, J., Niiniluoto, I. and Saarinen, E. (*eds.*) [1978]: *Essays on Mathematical and Philosophical Logic.* Dordrecht: D. Reidel Publishing Co. D.Fl. 95. Pp. 462.

Hirst, P. Q. [1979]: *Durkheim, Bernard and Epistemology.* Henley: Routledge and Kegan Paul. Paperback: £2.95. Pp. 212.

Hochberg, H. [1978]: *Thought, Fact and Reference—The Origins and Ontology of Logical Atomism.* Minneapolis: University of Minnesota Press. Pp. 489.

Hodges, W. [1977]: *Logic.* Middlesex: Penguin Books. Paperback: £1.25. Pp. 331.

Hodgkinson, C. [1978]: *Towards a Philosophy of Administration.* Oxford: Basil Blackwell. £9.50. Pp. 244.

Hofstadter, D. [1979]: *Gödel, Escher, Bach: An Eternal Golden Braid.* Hassocks: Harvester Press. £10.50. Pp. 777.

Home, R. W. and Connor, P. J. [1979]: *Aepinus's Essay on the Theory of Electricity and Magnetism.* New Jersey: Princeton University Press. £23.50. Pp. 514.

Hooker, C. A., Leach, J. J. and McClennen, E. F. (*eds.*) [1977]: *Foundations and Applications of Decision Theory. Volume I Theoretical Foundations,* Pp. 648. *Volume II Epistemic and Social Application,* Pp. 442. Dordrecht: D. Reidel Publishing Co. $72.00 complete set of 2.

Hooker, C. A. (*ed.*) [1979]: *The Logico-Algebraic Approach to Quantum Mechanics, II.* Dordrecht: D. Reidel Publishing Co. Paperback: D.Fl. 55. Pp. 466.

Hooker, C. A. (*ed.*) [1979]: *Physical Theory as Logico-Operational Structure.* Dordrecht: D. Reidel Publishing Co. D.Fl. 95. Pp. 334.

Høyrup, E. [1979]: *Books About Mathematics—A Bibliography.* Roskilde: Roskilde University Center. Pp. 180.

Huff, D. and Prewett, O. (*eds.*) [1979]: *The Nature of the Physical Universe.* Chichester: J. Wiley and Sons. £12.75. Pp. 140.

Hutchison, T. W. [1978]: *On Revolutions and Progress in Economic Knowledge.* Cambridge: Cambridge University Press. £12.50. Pp. 349.

Judson, H. F. [1979]: *The Eighth Day of Creation.* London: Jonathan Cape. £7.95. Pp. 689.

Kannegiesser, H. [1978]: *Knowledge and Science.* London: Macmillan Press. £6.95. Pp. 126.

Kaufmann, F. [1978]: *The Infinite in Mathematics.* Dordrecht: D. Reidel Publishing Co. D.Fl. 85. Pp. 235.

King, L. S. [1978]: *The Philosophy of Medicine—The Early 18th Century.* London: Harvard University Press. £12.25. Pp. 291.

Kneller, G. F. [1978]: *Science as a Human Endeavor.* New York: Columbia University Press. $17.45. Pp. 333.

Koenigsberger, D. [1979]: *Renaissance Man and Creative Thinking.* Hassocks: Harvester Press. £13.50. Pp. 282.

Koyre, A. [1978]: *Galileo Studies.* Hassocks: Harvester Press. £18.50. Pp. 278.

Kraupl Taylor, F. [1979]: *The Concepts of Illness, Disease and Morbus.* Cambridge: Cambridge University Press. £6.50. Pp. 131.

Krohn, W., Layton, E. T. and Weingart, P. (*eds.*) [1978]: *The Dynamics of Science and Technology.* Dordrecht: D. Reidel Publishing Co. Hardback: $29.50. Paperback: $14.95. Pp. 300.

Kuhn, T. [1978]: *The Essential Tension.* London: University of Chicago Press. £12.95. Pp. 366.

Kuhn, T. [1979]: *Black-Body Theory and the Quantum Discontinuity.* Oxford: Oxford University Press. £12.50. Pp. 356.

Kuipers, T. [1978]: *Studies in Inductive Probability and Rational Expectation.* Dordrecht: Reidel Publishing Co. $22.50. Pp. 142.

Lakatos, I. [1978]: *The Methodology of Scientific Research Programmes: Philosophical Papers, Vol. 1*; edited by John Worrall and Gregory Currie. Cambridge: Cambridge University Press. £9.00. Pp. 250.

Lakatos, I. [1978]: *Mathematics, Science and Epistemology: Philosophical Papers, Vol. 2*; edited by John Worrall and Gregory Currie. Cambridge: Cambridge University Press. £10.50. Pp. 285.

Larson, D. [1979]: *Nothing But Motion.* Oregon: North Pacific Publishers. $9.50. Pp. 292.

Lehman, H. [1979]: *Introduction to the Philosophy of Mathematics.* Oxford: Basil Blackwell. £8.50. Pp. 177.

Leinfellner, E. Z. (*ed.*) [1978]: *Wittgenstein and His Impact on Contemporary Thought.* Dordrecht: D. Reidel Publishing Co. Paperback: D.Fl. 80. Pp. 550.

Loggen, K. L. [1979]: *Aus dem Steinbruch.* Heusenstamm: Blick and Bild Verlag. Pp. 81.

Loux, M. J. [1979]: *Substance and Attribute.* Dordrecht: D. Reidel Publishing Co. D.Fl. 60. Pp. 187.

Lucchetta, G. [1978]: *Una Fisica Senza Matematica: Democrito, Aristotele, Filopono* Trento: Verifiche. L. 8,000. Pp. 188.

Ludwig, J. (*ed.*) [1978]: *Philosophy and Parapsychology*. New York: Prometheus Books. Hardback: $16.96. Paperback: $8.95. Pp. 454.

McMullin, E. [1978]: *Newton on Matter and Activity*. London: American University Publishers Group. £2.80. Pp. 160.

Macedo, J. [1979]: *Theoretical Basis of the Living System*. Missouri: Warren H. Green, Inc. $10.50. Pp. 107.

Magee, B. (*ed.*) [1978]: *Men of Ideas*. London: BBC Publications. £7.50. Pp. 312.

Mackay, D. [1978]: *Science, Chance and Providence*. Oxford: Oxford University Press. £3.50. Pp. 67.

Machamer, P. and Turnbull, R. (*eds.*) [1978]: *Studies in Perception*. Ohio: Ohio State University Press. $30.00. Pp. 567.

Machlup, F. [1978]: *Methodology of Economics and Other Social Sciences*. New York: Academic Press Inc. $25.00. Pp. 567.

Mandrou, R. [1979]: *From Humanism to Science 1480–1700*. Hassocks: Harvester Press. £11.50. Pp. 329.

Margenau, H. [1978]: *Physics and Philosophy: Selected Essays*. Dordrecht: D. Reidel Publishing Co. D.Fl. 75. Pp. 404.

Malatesta, M. [1978]: *Logistica*, Vol. 2. Roma: Libreria Editrice Rendenzione. Pp. 172.

Mattesich, R. [1977]: *Instrumental Reasoning and Systems Methodology*. Dordrecht: D. Reidel Publishing Co. £35. Pp. 352.

Medina, A. [1979]: *Reflection, Time and the Novel: Towards a Communicative Theory of Literature*. Oxford: Routledge and Kegan Paul. £7.95. Pp. 143.

Menger, K. [1979]: *Selected Papers in Logic and Foundations Diadactics, Economics*. Dordrecht: D. Reidel Publishing Co. D.Fl. 110. Pp. 341.

Meyer, M. [1979]: *Decouverte et Justification en Science*. Paris: Klincksieck. F. 94. Pp. 365.

Midgley, M. [1979]: *Beast and Man: The Roots of Human Nature*. Hassocks: Harvester Press. £7.50. Pp. 369.

Miller, J. W. [1978]: *The Paradox of Cause and Other Essays*. New York: W. W. Norton and Co. $14.95. Pp. 192.

Mittelstaedt, P. [1979]: *Quantum Logic*. Dordrecht: D. Reidel Publishing Co. D.Fl. 60. Pp. 149.

Moore, J. [1979]: *The Post-Darwinian Controversies*. Cambridge: Cambridge University Press. £18.00. Pp. 502.

Mulder, H. and Van de Velde-schlick (*eds.*) [1978]: *Moritz Schlick: Philosophical Papers, Vol. 1, 1909–22*. Dordrecht: D. Reidel Publishing Co. Hardback: D.Fl. 120. Paperback: D.Fl. 60. Pp. 365.

Nagel, E. [1979]: *Teleology Revisited*. New York: Columbia University Press. $20. Pp. 352.

Neill, W. T. [1978]: *Archeology and a Science of Man*. New York: Columbia University Press. $25.00. Pp. 321.

Oakeshott, M. [1978]: *Experience and Its Modes*. Cambridge: Cambridge University Press. £11.50. Pp. 359.

Orenstein, A. [1979]: *Existence and the Particular Quantifier*. Philadelphia: Temple University Press. $15.00. Pp. 185.

Ossorio, P. G. [1978]: *What Actually Happens*. South Carolina: University of South Carolina Press. Pp. 216.

Otero, M. H. [1977]: *La Filosofia de la Ciencia Hoy: Dos Approximaciones*. Mexico: Universidad Nacional Autonoma de Mexico. Pp. 73.

Parthey, H. (*ed.*) [1978]: *Problem und Methode in der Forschung*. Berlin: Akademie Verlag. DM 18.00. Pp. 246.

Patterson, C. [1978]: *Evolution*. Henley: Routledge and Kegan Paul. Hardback: £5.95. Paperback: £2.95. Pp. 197.

Peacocke, C. [1979]: *Holistic Explanation: Action, Space, Interpretation*. Oxford: Clarendon Press. £9.00. Pp. 219.

Platts, M. [1979]: *Ways of Meaning: An Introduction to a Philosophy of Language*. Henley: Routledge and Kegan Paul. Hardback: £7.50. Paperback: £3.95. Pp. 272.

Praetorius, N. [1978]: *Subject and Object*. Onsgardsvej: Dansk Psykologisk Forlag. Dkr. 25,00. Pp. 125.

Price, C. [1978]: *Landscape Economics*. London: Macmillan Press. £12.00. Pp. 168.

Putnam, H. [1979]: *Mathematics, Matter and Method Vol. I.* Second Edition. Cambridge: Cambridge University Press. Hardback: £12.50. Paperback: £4.95. Pp. 363.

Qvarnstrom, B. O. [1979]: *Formalizations of Trubetzkoy's Phonology.* Abo: Abo Akademi Forkningsinstitut. Pp. 191.

Rachie, K. *et al.* [1979]: *Tropical Legumes: Resources for the Future.* Washington: National Academy of Sciences. Pp. 331.

Radnitzky, G. and Andersson, G. (*eds.*) [1978]: *Progress and Rationality in Science.* Dordrecht: D. Reidel Publishing Co. Hardback: D.Fl. 95. Paperback: D.Fl. 40. Pp. 416.

Ree, J., Ayers, M. and Westoby, A. [1978]: *Philosophy and Its Past.* Hassocks: Harvester Press. Hardback: £8.50. Paperback: £3.50. Pp. 110.

Reichenbach, M. and Cohen, R. S. (*eds.*) [1978]: *Hans Reichenbach Selected Writings 1909–53*, Vol. 1. Dordrecht: D. Reidel Publishing Co. Hardback: D.Fl. 145. Paperback: D.Fl. 60. Pp. 501.

Reichenbach, M. and Cohen, R. S. (*eds.*) [1978]: *Hans Reichenbach Selected Writings 1909–53*, Vol. 2. Dordrecht: D. Reidel Publishing Co. Hardback: D.Fl. 130. Paperback: D.Fl. 60. Pp. 435.

Rescher, N. [1978]: *Scientific Progress: A Philosophical Essay on the Economics of Research in Natural Science.* Pittsburgh: University of Pittsburgh Press. $18.95. Pp. 277.

Rescher, N. [1978]: *Peirce's Philosophy of Science.* London: American University Publishers Group. £2.80. Pp. 125.

Rescher, N. [1979]: *Leibniz: An Introduction to His Philosophy.* Oxford: Basil Blackwell. Paperback: £7.96. Pp. 167.

Rescher, N. [1979]: *Cognitive Systemization.* Oxford: Basil Blackwell. £9.50. Pp. 211.

Roberts, G. W. (*ed.*) [1979]: *Bertrand Russell Memorial Volume.* London: George Allen and Unwin. £20.00. Pp. 488.

Ronan, C. A. and Needham, J. [1978]: *The Shorter Science and Civilization in China: 1.* Cambridge: Cambridge University Press. £7.95. Pp. 326.

Aldo Rossi, P. (*ed.*) [1979]: *Cibernetica e teoria dell'Informazione.* Brescia: Editrice la Scuola. L. 4,000. Pp. 244.

Rundle, B. [1979]: *Grammar in Philosophy.* Oxford: Clarendon Press. £14.00. Pp. 491.

Ruse, M. [1979]: *Sociobiology: Sense or Nonsense?* Dordrecht: D. Reidel Publishing Co. Pp. 231.

Saarinen, E. (*ed.*) [1978]: *Game-Theoretical Semantics.* Dordrecht: D. Reidel Publishing Co. D.Fl. 90. Pp. 383.

Saarinen, E., Hilpinen, R., Niiniluoto, I. and Provence Hintikka, M. (*eds.*) [1978]: *Essays in Honour of Jaakko Hintikka—on The Occasion of his Fiftieth Birthday January 12, 1979.* Dordrecht: D. Reidel Publishing Co. D.Fl. 80. Pp. 378.

Salmon, W. (*ed.*) [1979]: *Hans Reichenbach: Logical Empiricist.* Dordrecht: D. Reidel Publishing Co. D.Fl. 75. Pp. 750.

Sampson, G. [1979]: *Liberty and Language.* London: Oxford University Press. £5.75. Pp. 251.

Savage, C. W. (*ed.*) [1978]: *Perception and Cognition Issues in the Foundation of Psychology.* Minnesota: University of Minnesota Press. $25.00. Pp. 502.

Scheurer, P. [1979]: *Revolutions de la Science et Permanence du Réel.* Paris: Presses Universitaires de France. Pp. 368.

Schlanger, J. [1979]: *Une Theorie du Savoir.* Paris: Librairie Philosophique J. Vrin. Pp. 188.

Schlanger, J. [1978]: *Objets Idéels.* Paris: Librarie Philosophique J. Vrin. Pp. 108.

Schmid, A. F. [1978]: *Une Philosophie de Savant—Henri Poincaré et la Logique Mathematique.* Paris: Francois Maspero. Pp. 173.

Schmid, M. [1979]: *Handlungs-rationalität—Kritik einer dogmatischen Handlungswissenschaft.* Munchen: Wilheim Fink Verlag. Pp. 168.

Seidenfeld, T. [1979]: *Philosophical Problems of Statistical Inference.* Dordrecht: D. Reidel Publishing Co. D.Fl. 70. Pp. 250.

Shackle, G. L. S. [1979]: *Imagination and the Nature of Choice.* Edinburgh: University Press. £6.00. Pp. 159.

Sher, G. E. (*ed.*) [1978]: *Marxist Humanism and Praxis.* New York: Prometheus Books. Hardback: $14.95. Pp. 183.

Sherwood Taylor, F. [1979]: *Illustrated History of Science*. London: Heinemann Educational Books Ltd. £2.50. Pp. 178.

Shoesmith, D. J. and Smiley, T. J. [1978]: *Multiple-Conclusion Logic*. Cambridge: Cambridge University Press. £15.00. Pp. 396.

Sloman, A. [1978]: *The Computer Revolution in Philosophy*. Hassocks: Harvester Press. Hardback: £12.50. Paperback: £5.50. Pp. 304.

Spector, M. [1978]: *Concepts of Reduction in Physical Science*. Pennsylvania: Temple University Press. Hardback: $15.00. Paperback: $10.00. Pp. 114.

Spicker, S. F. (*ed.*) [1978]: *Philosophy and Medicine, Vol. 7 Organism, Medicine, and Metaphysics*. Dordrecht: D. Reidel Publishing Co. D.Fl. 65. $24.50. Pp. 326.

Spinner, H. [1978]: *Popper und die Politik—Vol. 1*. Berlin: Verlag J. H. W. Dietz Nachf. GmbH. Pp. 624.

Steffens, H. J. [1979]: *James Prescott Joule and the Concept of Energy*. Folkestone: Dawson Publishing Co. £10.00. Pp. 172.

Stent, G. S. [1978]: *Paradoxes of Progress*. San Francisco: W. H. Freeman and Co. Pp. 231.

Szabo, A. [1978]: *The Beginnings of Greek Mathematics*. Dordrecht: D. Reidel Publishing Co. D.Fl. 100. Pp. 353.

Tennant, N. W. [1978]: *Natural Logic*. Edinburgh: Edinburgh University Press. £6.00. Pp. 196.

Teyler, T. J. (*ed.*) [1978]: *Brain and Learning*. Dordrecht: D. Reidel Publishing Co. D.Fl. 50.00. Pp. 163.

Thomson, T. R. [1978]: *An Introduction to Cosmology*. Malmesbury: Taylor and Sons. £1.00. Pp. 24.

Thorpe, W. H. [1979]: *The Origins and Rise of Ethology*. London: Heinemann Educational Books. £5.50. Pp. 174.

Torretti, R. [1978]: *Philosophy of Geometry from Riemann to Poincaré*. Dordrecht: D. Reidel Publishing Co. D.Fl. 105. Pp. 459.

Tribe, K. [1978]: *Land, Labour and Economic Discourse*. Henley: Routledge and Kegan Paul. Hardback: £6.95. Paperback: £3.95. Pp. 182.

Turner, D. and Hazelett, R. (*eds.*) [1979]: *The Einstein Myth and the Ives Papers: A Counter Revolution in Physics*. Connecticut: Devin Adair Co. $22.50. Pp. 448.

Tymieniecka, A. T. (*ed.*) [1978]: *Analecta Husserliana: The Yearbook of Phenomenological Research Vol. VII—The Human Being in Action*. Dordrecht: D. Reidel Publishing Co. D.Fl. 80. Pp. 258.

Walford, G. [1979]: *Ideologies and Their Functions: A Study in Systematic Ideology*. London: George Walford. Hardback £3.95. Paperback: £1.95. Pp. 162.

Walker, R. C. S. [1978]: *Kant*. Henley: Routledge and Kegan Paul. £7.95. Pp. 201.

Weaver, H. [1978]: *Divining the Primary Sense—Unfamiliar Radiation in Nature, Art and Science*. Henley: Routledge and Kegan Paul. £4.75. Pp. 140.

Wechsler, J. (*ed.*) [1978]: *On Aesthetics in Science*. Cambridge, Massachusetts: MIT Press. £8.75. Pp. 180.

Weinberger, C. and Weinberger, O. [1979]: *Logik, Semantik, Hermeneutik*. Munchen: C. H. Beck'sche Verlagsbuchhandlung. DM 34. Pp. 231.

Williams, B. [1978]: *Descartes: The Project of Pure Enquiry*. Hassocks: Harvester Press. £10.50. Pp. 320.

Wilson, E. O. [1978]: *On Human Nature*. London: Harvard University Press. £7.85. Pp. 260.

Wilson, M. D. [1979]: *Descartes*. Henley: Routledge and Kegan Paul. £7.95. Pp. 255.

Winner, L. [1977]: *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. Cambridge, Massachusetts: MIT Press. $6.95. Pp. 386.

Wittgenstein, L. [1978]: *Remarks on the Foundations of Mathematics*. Oxford: Basil Blackwell. £15.00. Pp. 444.

Wittgenstein, L. [1979]: *Remarks on Frazer's 'Golden Bough'*. Nottingham: The Brynmill Press. £3.00. Pp. 42.

Wunderlich, D. [1979]: *Foundation of Linguistics*. Cambridge: Cambridge University Press. Hardback: £21.00. Paperback: £6.50. Pp. 360.

Yearly, L. H. [1978]: *The Ideas of Newman*. London: American University Publishers Group Ltd. £8.75. Pp. 188.

YOVEL, Y. (*ed.*) [1979]: *Philosophy of History and Action*. Dordrecht: D. Reidel Publishing Co. D.Fl. 75.00. Pp. 244.

ZIMAN, J. [1978]: *Reliable Knowledge—An Exploration of the Grounds for Belief in Science*. Cambridge: Cambridge University Press. £7.95. Pp. 197.

FIFTH INTERNATIONAL WITTGENSTEIN SYMPOSIUM

From 25 to 31 August 1980, in Kirchberg/Wechsel (Lower Austria, near Vienna). The theme of this year's symposium will be "Ethics: Foundations, Problems, and Applications".

The following sections are planned:

1. Wittgenstein, including a Seminar on Kant and Wittgenstein
2. Ethics, Deontic Logic, and Metaethics
3. Ethics and Social Sciences: Action and Decision Theory
4. Ethics and Justice
5. Medical Ethics, Ecological Ethics, Ethics and Religion

Scholars wishing to present a paper on a topic which falls within the above themes (papers for sections 2–5 need not be connected with Wittgenstein's philosophy) should request an abstract style sheet from either the President of the Austrian Wittgenstein Society, Dr A. Hübner, Markt 234, A-2880 Kirchberg am Wechsel/Austria, or Professor Werner Leinfellner, University of Nebraska, Department of Philosophy, Lincoln, Nebraska 68508/USA. The deadline for receipt of abstracts is 15 June 1980. In the case of too many contributions, the selection of papers will be subject to the decision of the organising committee. Contributors will be informed of the committee decision not later than 15 July. Official languages will be English and German.

Those planning to attend the conference should contact Dr Hübner at the above address for further information. The final date for registration is 15 June 1980. Conference fees will be 400 Austrian Schillings for non-students and 100 Austrian Schillings for students.

BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

ANNUAL CONFERENCE

The Society's 1980 Conference will be held at Chelsea College, University of London on 19–21 September. The following sessions have been arranged:

The Role of Mathematics in Physics

Main Speaker: Professor C. S. Sharma (Birkbeck)
Discussant: Professor I. W. Roxburgh (Queen Mary College)
Chairman: Professor H. R. Post (Chelsea)

What Makes a Good Experiment?

Main Speaker: Professor Sir Brian Pippard, FRS (Cambridge)
Discussant: Professor A. Franklin (Colorado)
Chairman: Professor Sir Harrie Massey, FRS (UCL)

Category Theory as a Foundation for Mathematics

Main Speaker: Dr J. L. Bell (LSE)
Discussant: Dr J. Mayberry (Bristol)
Chairman: Professor D. Scott, FBA (Oxford)

Relational Concepts of Space and Time

Main Speaker: Dr J. Barbour (Pavia)
Discussant: Dr W. Newton-Smith (Oxford)
Chairman: Professor C. W. Kilmister (Kings', London)

The conference is being organised on behalf of the Society by Dr D. A. Gillies and Dr M. L. G. Redhead, both of the Department of History and Philosophy of Science, Chelsea College, University of London, Manresa Road, London SW3 6LX, from whom further details may be obtained. (Please mark envelopes 'BSPS Conference'.)

MIND

Vol LXXXIX No 354 April 1980

A quarterly review of philosophy
edited by D.W. Hamlyn with the assistance of S.D. Guttenplan

Articles:
Knowledge, Luck and Charity: B. GRANT
Natural Kinds and Nominal Kinds: S. P. SCHWARTZ
Is Relevance Necessary for Validity?: G. ISEMINGER
Force and Sense: D. ZIMMERMAN
Frege on Thoughts: G. CURRIE

Discussions
Trust as a Problem for Utilitarianism: M. FISHER
Compulsive Behaviour and Sympathetic Concern: P. MEW
More on Mind-Body Problem and Indeterminacy of Translation: M. C. BRADLEY
Talk about Knowing: A. C. LLOYD
The Obstinance of Evil: F. CHERNOFF
Presupposition and Theories of Meaning: L. HOLLINGS

Book Reviews
Books Received
Orders to: Journals Department, Basil Blackwell Publisher Ltd.,
108 Cowley Road, Oxford OX4 1JF, England
Annual subscription:
Institutions: £10.00 or $22.50 (Canada $26.00) *Individuals:* £8.00 or $18.00
(Canada $21.00) *Single issues:* £3.00 or $6.75 (Canada $7.75)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

REVUE INTERNATIONALE DE PHILOSOPHIE

Editor : Michel MEYER
143, av. A. Buyl, 1050 Brussels, Belgium

Each number is devoted to a particular movement, a particular philosopher, or a particular problem.

Publishes 4 numbers annually or 1 double number and 2 other issues ; in English, French, German, or Italian.

Coming numbers for 1980 :

Philosophy of Science : contemporary issues
Aristotle
Scandinavian Philosophy

Our last issue was devoted to *Frege*

(articles by C. Imbert, E. D. Klemke, F. Jacques, J. Hintikka, R. Sternfeld, J. Bouveresse, T. C. Potts, M. Dummett, N. Mouloud, K. Nielsen, M. Loi)

RATIO

edited by STEPHAN KÖRNER

Volume XXI *December 1979* *No. 2*

CONTENTS

RUDOLF HALLER	*The Possibility of the Theory of Knowledge*
RODERICK M. CHISHOLM	*Socratic Method and the Theory of Knowledge*
PAUL LORENZEN	*The Philosophy of Science and Nelson's Theory of Knowledge as Illustrated on the Example of Geometry amd Ethics*
WILLIAM FRANKENA	*Methods of Ethics, 1977*
ANTHONY C. LLOYD	*Was Aristotle's Theory of Perception Lockean?*
GÜNTHER GREWEN-DORF	*Is Wittgenstein's Private Language Argument Trivial?*
HANS-ULRICHE HOCHE	*Does Goodman's 'Grue' Serve its Purpose?*

Annual Subscription: £12.50 (inland) £15.00 (overseas) $32.50 (USA and Canada)
Orders to: Journals Department, Basil Blackwell Publisher Ltd.,
108 Cowley Road, Oxford OX4 1JF, England.

The German edition of *Ratio* is published by FELIX MEINER VERLAG,
Richardstrasse 47, 2000 Hamburg 76, West Germany.

Published by BASIL BLACKWELL, Oxford

NOÛS

Nihil philosophicum a nobis alienum putamus

Hector-Neri Castañeda, *editor*

Volume XIII, Number 3 (September 1979)

Francis Jeffry Pelletier, *Sameness and Referential Opacity in Aristotle;* Clifton McIntosh, *Skolem's Criticisms of Set Theory;* Harry Silverstein, *Utilitarianism and Group Coordination;* Robert Young, *Compatibilism and Conditioning;* Eric W. Watkins, *Point of View in Depictive Representation. Reviews by* Chisholm, Humberstone, *and* Benardete.

Volume XIII, Number 4 (November 1979)

Special Issue on Counterfactuals and Laws

Nancy Cartwright, *Causal Laws and Effective Strategies;* Peter Van Inwagen, *Laws and Counterfactuals;* David Lewis, *Counterfactual Dependence and Time's Arrow;* G. Lee Bowie, *The Similarity Approach to Counterfactuals: Some Problems;* Joan Weiner, *Counterfactual Conundrum;* Steven De Haven & John King-Farlow, *Metaphilosophy and Religious Disagreements. Reviews by* Dworkin, Fogelin, *and* Kraut.

The March issue contains major papers to be read at the Spring meetings of the Western Division of the American Philosophical Association. For *Individuals* the subscription rates are: one year, $15.00; two years, $28.00; three years, $41.00. For *bona fide students*: $13.00. For *Institutions*: $25.00 per year (plus 10% for agency subscriptions). Add $1.00 per year for all non-U.S. subscriptions. Please make checks for one-year subscriptions payable to INDIANA UNIVERSITY-NOÛS. Checks for multiple-year and supporting subscriptions should be made payable to INDIANA UNIVERSITY FOUNDATION-NOÛS. Orders for subscriptions and changes of address should be sent to

The Secretary, NOÛS Department of Philosophy
Indiana University Bloomington, IN 47405

EDITORIAL BOARD

R. B. Braithwaite
G. Buchdahl
L. J. Cohen
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
W. C. Kneale
S. Körner
J. L. Mackie
D. J. O'Connor
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow
J. H. Woodger

EDITORIAL ADVISORY PANEL

The Editors are helped by members of the Editorial Board, members of the Committee of the British Society for the Philosophy of Science and by an Advisory Panel which includes:

Peter Clark
Jon Dorling
Arthur Fine
Michael Gardner
Donald Gillies
Ian Hacking
Colin Howson
Clive Kilmister
Noretta Koertge
David Miller
Alan Musgrave
Vernon Pratt
Ian Roxburgh
Geoffrey Sampson
Peter Urbach
Elie Zahar

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1980-81

President: Mr G. Buchdahl
President-Elect: Professor C. W. Kilmister
Hon. Secretary: Mr G. Hunt
Hon. Treasurer: Mr J. Powers
Members: Dr D. A. Gillies, Dr C. Phillips, Dr M. L. G. Redhead, Professor Y. Wilks

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all interested persons on payment of the annual subscription. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £2.50 for U.K. residents and £1.25 for overseas residents ($2.50 if resident in U.S.A. or Canada). Those who wish to join should write to the Hon. Secretary, Mr G. Hunt, Department of Philosophy, University of Warwick, Coventry CV4 7AL.

Brit. J. Phil. Sci. Vol. 31 No. 2, June 1980

The British Journal for the Philosophy of Science

25 FEB 1981

Contents

Articles

111–129 KATHLEEN WILKES Brain States

131–144 P. M. WILLIAMS Bayesian Conditionalisation and the Principle of Minimum Information

145–163 MICHAEL REDHEAD Models in Physics

165–173 R. FRANCIS On the Interpretation and Transitivity of Non-standard Synchronisms

Discussion

175–180 L. JONATHAN COHEN Some Comments on Third World Epistemology

Review Article

181–192 THOMAS KUHN The Halt and the Blind: Philosophy and History of Science (review of C. Howson (*ed.*): *Method and Appraisal in the Physical Sciences*)

Reviews

193–203 NORETTA KOERTGE on Gerald Holton: *The Scientific Imagination: Case Studies*; C. B. SCHMITT on R. E. Butts and J. C. Pitt (*eds.*): *New Perspectives on Galileo*; NIGEL HOWARD on E. Ullman-Margalit: *The Emergence of Norms.*

Recent Publications

Printed in Great Britain at Aberdeen University Press

Volume 31 Number 3 September 1980
BJPIA5 31 (No. 3) 213-316 (1980) ISSN 0007-0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: JOHN WORRALL

Subscriptions

The *Journal* is published quarterly. The subscription price for mailing addresses in Britain is £7.00 net per volume post free and for mailing addresses outside Britain £11.00 (or foreign currency equivalent calculated at the day's rate); single issues £2.50 net (overseas £4.00) plus postage. Orders may be sent to any bookseller or direct to Aberdeen University. Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

For members of the British Society for the Philosophy of Science—who should order direct from Aberdeen University Press—the subscription price is £4.50 (overseas £7.00). Membership of the Society is international and is by no means confined to people resident in the United Kingdom.

Microform

The current volume of the *Journal* is available in microform to subscribers to the printed version, *at a combined print/microfiche rate of* £12.50 *(overseas* £16.50).

Back Volumes

Volumes 1–16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17–30 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application. Volumes 1–29 are also available in microform from Kraus Reprint Corporation.

Typescripts and Review Books

Typescripts, correspondence and books for review should be sent to the Editor at the Department of Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street, London WC2A 2AE.

Review books are accepted on the understanding that, although they will appear in the 'Recent Publications' list, a full review cannot be guaranteed, and that books not reviewed are not returnable.

Notes to Contributors

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Detailed instructions on the lay-out of typescripts are published occasionally at the end of the *Journal* and anyway may be obtained from the Editor.

Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

© British Society for the Philosophy of Science 1980

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editor.

Brit. J. Phil. Sci. 31 (1980), 213–254 *Printed in Great Britain*

A New *Begriffsschrift* (I)

by JOHN P. MAYBERRY

1 *Introduction*
  (*i*) *The Purpose of the Conceptual Notation*
  (*ii*) *Frege's Dogmatism*
2 *Frege's Principal Doctrines*
  (*i*) *Functions and Objects*
  (*ii*) *Sense and Reference*
  (*iii*) *Logical Objects: Truth Values and Value-Ranges*
  (*iv*) *The Scope of the Conceptual Notation*
3 *The Notion of Set*
  (*i*) *The Notion of Set and the Notion of Number*
  (*ii*) *Cantor's Doctrine of Infinity*
  (*iii*) *The Logical and Ontological Consequences of Cantor's Doctrine*
4 *Exposition of the new* Begriffsschrift
5 *Conclusions*

1 INTRODUCTION

It is now one hundred years since Frege inaugurated the modern, mathematical era of logic with the publication of his *Begriffsschrift*. The importance of his technical innovations—in particular his invention of the quantifiers—is now quite beyond dispute. Moreover, his ideas on the nature of language are well-known and widely discussed, even though they have not achieved the universal acceptance accorded his fundamental logical doctrines. But his ideas on the foundations of mathematics have fallen completely into disfavour. Not only his positive doctrines, but even his brilliant attacks on rival views, are simply ignored.

This neglect is easier to explain than to justify. It stems, no doubt, from the opinion that the formal inconsistency of his system has discredited the philosophical ideas which underlie it. I believe, however, that this opinion is profoundly mistaken. Frege's basic doctrines are of central importance, even—no, especially—for finitary or constructive theories of the foundations of mathematics. This, in any event, is my principal thesis. I shall present my case by expounding a theory based on Fregean doctrines, a 'new

Received 23 *July* 1979

Begriffsschrift' which corrects the errors of Frege's original theory, and which is sufficiently flexible to incorporate either the classical, or the finitary, point of view.

But before I begin the exposition of my theory, or even a detailed analysis of Frege's principal doctrines, I want to touch briefly on two aspects of Frege's approach to the foundations of mathematics which may give rise to certain misconceptions. I fear that these misconceptions may prove to be obstacles to a fair assessment either of Frege's theory or my own, for they concern the very nature of our common enterprise.

1(*i*) THE PURPOSE OF THE CONCEPTUAL NOTATION

Throughout his entire career, Frege's primary interest lay in the foundations of mathematics, and, in particular, of arithmetic. Everything else—his speculations on the nature of language, even his logic—was secondary. He tells us in the preface to the *Begriffsschrift* that he developed his conceptual notation as a tool which would enable him to give a logically precise definition of the concept of 'ordering in a series', the concept in terms of which he planned to give an account of number. Thus from the very beginning Frege saw his conceptual notation as an instrument—a means to his ends, not an end in itself. In particular, he was not engaged in analysing the logic of ordinary language; it was rather his intention to develop an artificial substitute for language, or more exactly, for language as employed for certain very restricted ends, namely, mathematical ones. The conceptual notation, he tells us, is to language what the microscope is to the eye.

Indeed, Frege goes out of his way to emphasise the narrow aims which his conceptual notation was designed to achieve. He takes care, for example, to distinguish his project from the vastly more ambitious enterprise of Leibniz:

> Leibniz also recognized—perhaps overestimated—the advantages of an adequate method of notation. His idea of a universal characteristic, a *calculus philosophicus* or *ratiocinator*, was too ambitious for the effort to realize it to go beyond the mere preparatory steps. The enthusiasm which overcomes its [would be] creator when he considers what an immense increase in the mental power of mankind would result from a method of notation which fits things themselves lets him underestimate the difficulty which such an undertaking confronts ([1879], p. 105).

Frege frequently calls attention to the contrast between the conventions governing 'language' and those governing his conceptual notation. And, indeed, his conceptual notation represents, in its grammar, a radical departure from real languages. It has no adjectives, adverbs, prepositions, or pronouns; no tenses; no common nouns (this is especially significant); no distinction between singular and plural forms, since all 'subjects' are, in any case, singular. Its treatment of generality by means of quantifiers is entirely different from that of actual languages.

Indeed, it is clear that the conceptual notation is a language only in a Pickwickian sense: you can't speak it, and what you can write in it is severely limited—you can't ask questions, make requests or promises (there are no pronouns or other devices for context dependent reference) nor employ metaphors. The notation was designed to exclude the possibility of expressing nonsense: 'Every proposition of *Begriffsschrift*' Frege tells us, 'expresses a thought'. Of course no real language could meet this last requirement. That would make the language too inflexible to meet the demands of everyday use, not to mention those of literary or scientific expression.

Perhaps it will be thought that I make too much of this matter. But it seems to me essential that we take the limited use intended for the conceptual notation into account, especially when we are discussing its semantics. Otherwise we shall tie ourselves in knots in the mistaken belief that our semantics must apply, not only to the conceptual notation, but to language itself. Such a misconception is not just an abstract possibility. Frege himself fell victim to it, as his peculiar claim that the concept *horse* is not a concept shows. We arrive at Frege's position if we suppose that the semantics of complete and incomplete signs in the conceptual notation extends to expressions like 'the concept *horse*' in actual languages. But why should we suppose this?

It is true that Frege was tempted, from time to time, to apply his semantical theory to language, as in his treatment of indirect discourse in 'Sense and Reference', for example. Some of these applications are very interesting. But they are all peripheral to his principal aim. In any case, we need not follow Frege in these speculations, since they have no bearing on the question of whether his semantical theory is adequate for the restricted purposes of the conceptual notation.

I do not mean to imply that when we are discussing the foundations of mathematics in 'ordinary language' (how else, indeed, could we discuss it?) we can be anything but scrupulous in what we say and in how we say it. But we are under the same obligation whatever the subject under discussion. And surely the suggestion that we need a theoretical semantics for natural language before we can undertake an investigation of those foundations is a counsel of despair. There are, to be sure, many snares and pitfalls in our path. For example, there are many notions which have unexceptionable instances of application, but whose range of applicability is uncertain. The notion of 'truth' is a notorious case in point. And there is no way that we can guarantee, in advance, that we shall not stray into error or nonsense when we employ these notions in ordinary discourse. But here the conceptual notation can come to our rescue. For we need not transfer such notions to our conceptual notation in their entirety. There we can confine ourselves to modes of expression which are less controversial and problematic. The conceptual notation represents, so to speak, the last ditch of defence, protecting only those forms of discourse which are essential to our enterprise.

I(*ii*) FREGE'S DOGMATISM

We are searching for a theory of the foundations of mathematics. But what kind of theory ought it to be? For Frege, the first task of such a theory is to tell us what we are talking *about* in mathematics, to explain what mathematical propositions mean.[1] The question of mathematical knowledge and its origins did not particularly interest him. Indeed, he felt this question to be most inopportune, if it is asked at the beginning of an enquiry into the foundations of the subject. Frege's approach thus represents a decisive break with the central tradition of modern philosophy since Descartes, a tradition in which such epistemological questions are regarded as primary.[2]

This brings us to that aspect of Frege's theory which is most likely to give rise to misunderstanding: I refer to his dogmatic style of exposition. But it is perhaps misleading to speak of 'style' here, for Frege's dogmatism lies not just in the manner of his exposition, but in its substance. Of course I am using the term 'dogmatic' in a descriptive, rather than in a pejorative, sense here. There may be those for whom this is a distinction without a difference. But they will, for that reason alone, be incapable of grasping what it was that Frege was trying to achieve.

Frege was self-consciously dogmatic in logic, not in the sense that he regarded his theory as immune to rational criticism, but in the sense that he held that the laws of logic cannot be justified by appeal to the propositions of any other science—certainly not those of psychology. The laws of logic must, in the final analysis, simply be laid down without proof, even when we have doubts concerning their validity, as he himself admitted from the beginning was the case in respect of his Basic Law V. This is not to say that we are at liberty to lay down these laws arbitrarily. For to claim that someting is a law *of logic* is to claim, at the very least, that it is universally valid and binding on our reasoning.

---

[1] And notice: an explanation of 'what' a proposition means is not to be confused with an explanation of 'how' it means. The latter is an altogether more formidable task.

[2] This has been pointed out by Dummett ([1973], ch. 19).

Since this is so important a matter, it is perhaps best that I allow Frege to speak for himself:

> The question why and with what right we acknowledge a law of logic to be true logic can answer only by reducing it to another law of logic. Where that is not possible logic can give no answer. If we step away from logic we may say: we are compelled to make judgements by our own nature and by external circumstances; and if we do so we cannot reject this law—of Identity, for example; we must acknowledge it unless we wish to reduce our thought to confusion and finally renounce all judgement whatever. *I shall neither dispute nor support this view.* I shall merely remark that what we have here is not a reason for something's being true, but for our taking it to be true. ([1893*a*], p. 15; emphasis supplied).

But to grasp the full significance of these logical views we must place them in the context of Frege's theory as a whole. And here we come to what is most characteristic of his approach to the foundations of mathematics, and what constitutes his single most important contribution to philosophy—indeed, it is what forms the basis of his claim to stand among philosophers of first rank. Frege's logic is an integral part of a general system of metaphysics. The simplicity and precision of his logic rests upon a prior metaphysical analysis of the categories of being, and a careful delineation of their boundaries. In Frege's theory, metaphysics is restored to its rightful place as 'first philosophy'. His theory thus represents a return to the point of view taken in classical antiquity and in the Middle Ages. But it is not simply a return to an earlier theory. On the contrary, it is entirely novel, and, moreover, vastly superior to its classical and medieval precursors in point of simplicity, explanatory power, and precision.

Frege's dogmatic treatment of logic extends to the whole of his system including the metaphysics. Indeed, it is impossible to draw a sharp line distinguishing his logic from his metaphysics. Each informs and supplements the other. And this is a strength of his theory, not a weakness. In any case it is a necessity. Logic and metaphysics are inseparable. And any logico-metaphysical theory must be dogmatic as regards its fundamental principles, a basic truth which Aristotle and his followers fully appreciated. Frege's originality among modern thinkers consists in his explicit and honest recognition of it, as witnessed by the passage I quoted above. Moreover, he shows us, in his polemical writings, the terrible confusion which arises when one looks outside metaphysics and logic for one's fundamental mathematical principles.

In Frege's system we encounter the first example of a metaphysical theory which is genuinely 'scientific' in the modern sense of that much abused word. It is not widely speculative; it is modest in its aims and scope;

it makes no attempt to solve the 'riddle of the world', and no claims to having done so; it has nothing to say on the subjects traditionally (and, in my view, wrongly) held to be the central problems of metaphysics—God, free will, the human soul, *etc.*; on the contrary, it deals with the simplest and most general features of things—with 'being as such' in fact.

But can a genuinely scientific theory be dogmatic in character? Well, astronomy can't be, nor can biology. The purported truths of these sciences must be brought before the tribunal of empirical observation. But mathematics does not deal with so specific and limited a subject matter. It is applicable to all the particular sciences and circumscribed by none. It cannot rest upon empirical observation, because all empirical observation, except the most utterly trivial, presupposes mathematics, not to mention logic. In any case, Frege's system is not 'dogmatic' in the pejorative sense. What is claimed on its behalf is not absolute and incontrovertable certainty, but truth. And any science must lay claim to truth, however tentatively and with whatever qualifications.

But it is pointless to pursue this line of argument. One cannot establish (or refute) the possibility of a scientific metaphysics by *a priori* argument. One can only judge the claims advanced on behalf of a particular theory which is expounded in sufficient detail. The new *Begriffsschrift* which I shall present is such a theory. It is based on Fregean principles, but incorporates notions—in particular, the notion of *set*—which have been developed by other thinkers—Cantor, Brouwer and Zermelo, to name the most important.

I have divided my exposition into three parts. The first (section 2) is a critical exposition of Frege's principal doctrines. The second (section 3) deals with the notion of set and Cantor's theory of the infinite. The third (section 4) consists of a systematic presentation of the new *Begriffsschrift*.

2 FREGE'S PRINCIPAL DOCTRINES

It is obvious that our new *Begriffsschrift* must differ from the old one, if only in order to avoid contradiction. But how much of the old theory should be preserved? Which of Frege's basic doctrines are essential to our purpose, which are merely peripheral, and which actually harmful? What, in short, can we, and ought we to, salvage from the wreckage of his system?

Frege's ideas changed considerably over the fourteen years between the publication of the *Begriffsschrift* in 1879 and the appearance, in 1893, of volume I of the *Grundgesetze*. In the introduction to the *Grundgesetze* he tells us that he was forced to discard an earlier manuscript in order to accommodate the developments in his logical views, developments to which he attached the greatest importance even though, as he ruefully admitted, they 'stamped those views with an impress of paradox'.

There is, of course, a certain overlap between the original and the mature versions of his system. For the logical grammar and syntax of the *Begriffsschrift* are retained in the *Grundgesetze*, though in a considerably developed form. But the semantics of the system, and the underlying ontology on which that semantics depends, are radically altered. Indeed, the earlier version of the theory can scarcely be said even to have a semantics in the proper sense. His ontology, which is primary in respect of logical priority, was the last part of his theory to be developed.

But why did Frege feel compelled to convert what began as a system of logic into a system of metaphysics? The answer, in brief, is that he came to believe it essential that logic and the foundations of mathematics be freed of any taint of psychology. The fundamental question, as he came to see it, is not 'How do we know?', but 'What are we talking about?' He had to explain, first of all, what kinds of things there are, and, secondly, how his conceptual notation could be used to refer to them. He needed, in short, an ontology, and a referential semantics based upon it.

Frege's attack on the 'corrupting incursion of psychology into logic' was not an attack on psychology *per se*. His quarrel was with those who took the view that psychology, whether empirical or transcendental, is prior to logic and mathematics. Such a view was, for Frege, a counsel of despair. By denying the claim of logic and mathematics to objectivity, it would lead to the destruction of both these disciplines, and with them, of all of objective science, including even psychology itself.

Frege's attack on 'psychologism' first appeared in the *Grundlagen*. And it is clear that much of his criticism there of the opinions of other thinkers could be applied to his own earlier views. At least, there is nothing he says in the *Begriffsschrift* which would rule out a psychologistic interpretation, and much which would appear to call for one. Perhaps there is something of the zeal of the repentant sinner and recent convert in Frege's vigorous assertion of his later views. In any case, when he came to write the *Grundgesetze* his aim was complete objectivity, undiluted by subjective psychological notions. The means to this aim were a clarified and simplified conceptual notation, a radically simple ontology, and a new semantical theory which explained how the one could be used to refer to the things classified by the other. Frege began by attempting to construct 'a formula language of pure thought modeled on the formula language of arithmetic'; in the end he produced a system of metaphysics. I shall now turn to a detailed examination of the principal doctrines of that system.

2(*i*) FUNCTIONS AND OBJECTS

For Frege, the distinction between functions and objects is the most important single doctrine underlying his mature system. It is a metaphysical distinction, but it originated in the grammatical distinction, drawn in the *Begriffsschrift*, between functions and arguments. Even this early version of the doctrine struck Frege as being of central importance:

> I hope that logicians, if they do not allow themselves to be frightened off by the first impression of unfamiliarity will not refuse their assent to the innovations to which I have been driven by a necessity inherent in the subject matter itself. These deviations from the traditional find their justification in the fact that logic up to now has always confined itself too closely to language and grammar. In particular, I believe that the replacement of the concepts of *subject* and *predicate* by *argument* and *function* will prove itself in the long run. ([1879], p. 107)

The importance he attached to the metaphysical version of this distinction can be gauged by the fact that he begins the exposition of the *Grundgesetze* with two sections devoted to explaining it. Indeed, the central idea which, for him, forms the basis for the distinction is actually given as the title to section §1: 'The function is unsaturated'.

To explain what this means, and to explain what sort of thing a function is, Frege invites us to consider what sort of expression can be used to denote functions. The expression

$$(2+3\cdot\xi^2)\cdot\xi$$

for example, denotes a function by virtue of the fact that we can obtain expressions denoting the various values of the function by substituting expressions denoting the corresponding arguments for the letter 'ξ' throughout the given expresssion. Thus on substituting the numeral '3' for 'ξ' we obtain the expression '$(2+3\cdot3^2)\cdot3$' which denotes the value 87. The occurrences of the letter 'ξ' in the expression denoting the function therefore serve merely to indicate where the substitution of expressions denoting its arguments can occur. Frege regards the letter 'ξ' as rendering visible, as it were, the 'gaps' in the function sign which require filling in order that the resulting sign should denote a value of the function. What designates the function itself is that part of the expression which remains when the letter 'ξ' is deleted, provided that all of the resulting gaps are recognized as requiring filling by instances of one and the same argument-denoting expression. The need for this latter proviso becomes apparent if we consider an expression which denotes a function of two arguments:

$$(2+3\cdot\xi^2)\cdot\zeta$$

Here the gap filling letters may be replaced by distinct argument designators, so the gaps are no longer linked. In this sense, then, the sign for a function is said to be 'incomplete'. And this incompleteness in the function sign corresponds, according to Frege, to that feature of the function itself by virtue of which it may be said to be 'unsaturated'.

Frege goes on to distinguish different *levels* of functions corresponding to different kinds of unsaturatedness. For example, in the expression

$$\frac{d}{dx}(f(x))\Big|_{x=0}$$

the letter 'f' plays a role analogous to that of 'ξ' in the earlier example, but it must be replaced by the name of a first level function of one argument if we are to obtain a name of a value. For example, if '$f(\xi)$' is replaced by '$(2+3\cdot\xi^2)\cdot\xi$' the resulting expression designates the value 2. The original expression is thus said to designate a function of the *second level.*

Frege completes his list of ontological categories by characterising *objects* as those entities which are designated by means of complete signs, like '2' and '3+5', *etc.*, and which do not, therefore, exhibit the unsaturatedness which distinguishes functions.

The problem for the student of Frege's ontology is to make sense of this fundamental metaphor of 'unsaturatedness' and to discern wherein lies the essential difference between 'saturated' objects and 'unsaturated' functions. But before we proceed any further, we should briefly take notice of the power and simplicity of the semantics to which this ontology gives rise.

All complete expressions in Frege's conceptual notation designate objects (or 'indefinitely indicate' them, in the case of expressions containing free variables). And this applies even to expressions which correspond to sentences in ordinary language. The latter designate one of the two truth values, the True or the False. These Frege takes to be fully fledged objects, existing in their own right. All of this means that the primary semantical function of complete expressions in the conceptual notation is that of reference. Indeed, this is true of *all* signs of the conceptual notation, complete or incomplete. The essential semantical function of each sign is to designate its reference. And this it does quite independently of the context in which it occurs. Thus Frege's well-known observation in the *Grundlagen* that 'only in the context of a sentence does a sign have reference' clearly does not apply to the conceptual notation of the *Grundgesetze*. Sentences have no distinguished semantical role to play in the latter system. The reference of a complex expression is obtained simply by combining the references of its constituent simple signs in exactly the order in which those signs themselves are combined in the given expression. Reference is thus a kind of 'morphism' or structure preserving map from the conceptual notation into the ontology. This is the sense in which we may describe Frege's theory as a 'referential semantics'.

In the conceptual notation of the *Grundgesetze* there is thus a perfect harmony between the syntactical properties of the signs and the ontological properties of the things they signify. Indeed, this is perhaps the most impressive single feature of Frege's system. But it raises the question to what extent the ontological distinction between objects and functions depends upon the grammatical distinction between complete and incomplete signs. Clearly the ontological distinction ought to be the prior one. But Frege seems to rest most of his case for that distinction on its grammatical parallel *via* the metaphor of 'saturation'. In the conceptual notation itself this causes no difficulty, because there function signs always occur only as constituent parts of complete expressions which denote objects. There is no provision made for denoting functions independently, for allowing their signs to occur entirely on their own as proper denoting expressions. The latter use of function signs occurs only in the informal exposition of the theory. But this informal use of expressions which denote functions independently leads to a curious difficulty.

The difficulty occasioned by the independent use of function signs is discussed in '*Uber Begriff und Gegenstand*'. There Frege defends the paradoxical claim that expressions like 'the concept horse', which appear to denote concepts, in fact denote objects:

> In logical discussions one quite often needs to assert something about a concept, and to express this in the form usual for such assertions—*viz*. to make what is asserted of the concept into the grammatical predicate. Consequently, one would expect that the reference of the grammatical subject would be the concept; but the concept as such cannot play this part, in view of its predicative nature; it must first be converted into an object, or, speaking more precisely, represented by an object. We designate this object by prefixing the words 'the concept', *e.g.*
>
> The concept *man* is not empty
>
> Here the first three words are to be regarded as a proper name which can no more be used predicatively than 'Berlin' or 'Vesuvius'. ([1892], p. 46)

But what applies to the special case of concepts also applies to functions in general, as he goes on to explain in a footnote:

> What I call here the predicative nature of the concept is just a special case of the need of supplementation, the 'unsaturatedness', that I gave as the essential feature of a function in my work '*Function und Begriff*' (Jena, 1891). It was there scarcely possible to avoid the expression 'the function $f(x)$', although there too the difficulty arose that what this expression stands for is not a function. ([1892], p. 47)

Surely we are entitled to ask here: what is the nature of these bizarre objects which serve as references for expressions like 'the concept man' or 'the function $f(x)$'? And what answer to this question could possibly satisfy us? It would appear that in Frege's functions we have a whole category of entities which we can't even talk about!

This peculiar difficulty arises as a direct consequence of Frege's taking the metaphor of 'unsaturatedness' as the sole basis for the function-object distinction.[1] The essential nature of functions is thus inextricably bound up with the grammatical role of function names in the conceptual notation. This, in turn, requires that the expressions used informally to designate functions independently must be employed with great circumspection. For example, the two expressions

$$`(\xi+1)^2\text{'} \quad \text{and} \quad `\xi^2+2\cdot\xi+1\text{'}$$

can, in certain contexts, be used directly to designate functions. But these contexts must be carefully circumscribed. Thus upon setting down the expression

$$`(\xi+1)^2 = \xi^2+2\cdot\xi+1\text{'}$$

we obtain, on Frege's own account, not an equality between functions, but merely a name of a concept, that is, a function all of whose values are truth values.

This last example shows why it is not possible according to Frege, to directly assert the identity of two functions. Of course, the reasons for this are to be found in the special conditions he lays down for the informal employment of function signs. I believe that those reasons are specious. But the prohibition of assertions of identity between functions is sound. Moreover, this prohibition constitutes the real basis for the necessity of sharply distinguishing functions from objects.

I maintain that the necessity for this distinction rests, not upon the grammatical properties of the signs which designate functions and objects, but upon the logical impossibility of formulating classical identity conditions for functions: functions are not properly individuated and therefore cannot be classified as objects. The satisfaction of clear and unambiguous classical identity conditions is the defining property of entities in the category of objects. The fact that they satisfy the familiar classical laws of identity serves to distinguish objects from all other entities, and there is no need to employ metaphors based on the grammatical properties of signs in order to establish this distinction. On the contrary, the grammatical differences between function signs and object signs rest upon the prior ontological distinction between functions and objects. There is no reason why we cannot use 'complete' signs to designate functions, no reason why we cannot talk directly about functions, as long as we do not try to apply classical identity to them.

---

[1] It also stems from his tendency to extend the semantics of the conceptual notation to natural language, as I pointed out in section **1**(*i*).

Of course I have not yet given the reasons why functions cannot be the subjects of assertions of indentity which obey the usual classical laws of logic. And I am not yet in a position to do so. For I have not yet discussed the all-important Cantorian doctrine of the 'absolute infinity' of the category of objects, or that of the fundamental distinction between the 'transfinite' and the 'absolutely infinite' on which the former doctrine depends. As I shall argue in section 3, these Cantorian doctrines, however unfamiliar, are really of fundamental importance for the whole of modern mathematics. Yet they are not to be found among the doctrines underlying Frege's system. This, I shall argue, is the principal cause of the disaster which overtook it. Nevertheless things can be put right. Frege's system can be modified to incorporate these doctrines, and thus to provide, in the projected new *Begriffsschrift*, a comprehensive foundation for mathematics of the sort he originally had in mind.

It will thus prove possible to characterise the difference between functions and objects without resorting, as Frege does, to a grammatical metaphor. Not that we need reject out of hand Frege's description of functions as 'incomplete' entities. For this description is, as we shall see, appropriate, though in a sense different from that which Frege intended. But we need not postulate, as he did, an object *Doppelgänger* for each function, which we refer to when we want to say something about the function itself.

There is, however, a more radical way of distinguishing functions and objects. This is simply to treat function signs as syncategorematic signs, to deny them their status as referring signs. To do things this way would obviously avoid altogether the question of what sort of entities function signs denote. Such an approach would have the extra attraction of yielding an ontology in which the notion of being would be 'univocal' in the scholastic sense: the only mode of existence would be that of objects. Of course the semantics required by this approach would lack the immediacy and simplicity of Frege's semantics, since there would no longer be a complete parallel between the grammatical and ontological categories of the theory. But other, and more serious, issues are at stake. It is not, after all, clear that functions, as real entities, can be dispensed with. For without functions, there are serious problems concerning the justification of definition by transfinite recursion. And this, in turn, complicates the theory of cardinal numbers. In particular, the definition of the sequence of Cantor's $\aleph$'s would be called into question. Thus it must remain a moot question whether this kind of 'nominalistic' alternative to Frege's theory is at all tenable. This is a question to which I shall later return when I come to discuss the new version of the *Begriffsschrift*. Nevertheless, however it may be resolved, we must credit Frege with having drawn attention to something of fundamental importance for the foundations of mathematics when he established the distinction between functions and objects and placed it at the centre of his theory.

2(*ii*) SENSE AND REFERENCE

I have already explained how Frege's semantics ought to be conceived as a semantics of reference. The notion of reference, then, is of central importance in Frege's system. But it is not a problematic notion; it does not call for an elaborate exegesis. Such difficulties as do arise in connection with reference are really concerned with the nature of the entities referred to, *i.e.* with ontology. Why then is so much fuss made over it? Simply because it is invariably discussed in connection with the altogether more problematic and dubious notion of *sense*. The relative unimportance of the notion of sense, for the conceptual notation, can be gauged by the fact that it is rarely mentioned in the *Grundgesetze*. Yet it has perhaps attracted more attention from scholars than any other of Frege's doctrines. How are we to attain a balanced and judicious assessment of the intrinsic importance of this notion and of its place in Frege's system?

Perhaps we should turn to Frege's own treatment of the subject. Frege introduced the sense-reference distinction in the course of thinking his way out of the hopelessly confused and incorrect account of identity which he gave in the *Begriffsschrift*. There he argued that a statement of identity

$$a = b$$

asserts a relation between the expression 'a' and the expression 'b' rather than between the objects denoted by those expressions. He even employed a three barred identity sign

$$a \equiv b$$

presumably in order to distinguish it from 'ordinary' identity, and translated the resulting expression by:

The symbols 'a' and 'b' have the same conceptual content.

He later came to regard this account as seriously defective. And so, of course, it is. When we make assertions of indentity, as in

$$e^{2\pi i} = 1$$

or

The morning star is identical with the evening star.

We do not intend merely to assert something about our conventions for designating complex numbers or astronomical bodies. We intend to say something about the objects themselves: something *objective*, in short.

This suggests the view that the identity sign designates a real relation applying to objects, not a linguistic relation applying to names. On the other hand, there is something rather odd about a relation in which each thing stands to itself only, and to no other thing. Or rather, there is something odd about the things we *say* concerning this relation. We expect objective discourse to tell us about the relations that obtain between things. But what does an identity statement tell us? That something is itself and nothing else? How are we, asks Frege, to account for the difference in "cognitive value" between the statements '$a = a$' and '$a = b$', assuming the latter to be true?

Well, the statements '$a = a$' and '$a = b$' do strike us as different. And it is perhaps natural that we should ask why this is so. But this question is only a particular case of the question why any two distinct statements strike us as different. The problem of *distinguishing* the 'cognitive values' of arbitrary statements leads us inevitably to the general problem of *determining* the 'cognitive value' of any statement whatsoever, *i.e.* of determining how it is that we come to understand statements in general. But surely this is a problem for psychology. No doubt Frege would maintain that by the 'cognitive value' of an expression he means something objective, something to be sharply distinguished from the idea in the mind which it evokes or to which it corresponds. In this way, perhaps, he could avoid the direct imputation of psychologism, but only at the expense of hypostatising a whole realm of entities—in fact, the realm of sense—which somehow mediates between our 'inner' mental world and the 'external' world of objects.

Frege's sense-reference distinction is thus directed to the solution of a problem that, properly understood, falls outside the province of logic. It 'explains' what we grasp when we understand an expression, what it is which allows us to pass from the expression to its reference. But grasping a sense is every bit as mysterious an affair as grasping a reference, more mysterious, in fact, if senses are regarded as peculiar sorts of objects, because, in some cases anyway, we have a clear idea of what the reference is which we grasp.

We should look on Frege's theory of sense, not as a metaphysical theory, but as a parenthetical remark, as it were, forestalling a misconception which might otherwise arise concerning the notion of reference. For the latter notion must not be confused with the vague, informal notion of the 'meaning' of an expression, a confusion especially likely to arise in German because of Frege's use of the word '*Bedeutung*' which is customarily translated as 'meaning'. We must distinguish the reference of an expression from the way in which that reference is determined, *i.e.* from its sense—a perfectly unexceptionable, if rather pedestrian, piece of advice. But we must not be tempted into regarding the expression 'the way in which the expression so and so determines its reference' as itself having a reference, that is to say, an object which it denotes: the sense of the expression so and so.

Did Frege himself commit this blunder? Judging on the basis of '*Über Sinn und Bedeutung*' we would have to say yes. But a more charitable construction can be placed on his brief remarks in the *Grundgesetze* (*e.g.* in §32 of volume I). In any case, as long as we do not regard senses as objects the notion is harmless, indeed, useful.

Notice that on this explanation the sense-reference distinction is still something perfectly objective. There is nothing psychological or subjective about the way in which the reference of an expression of the conceptual notation is determined: it depends upon the references of the constituent simple signs—which are determined either 'ostensively' (*i.e.* by direct stipulation), or by formal definitions—upon the manner in which those simple signs are combined, and, finally, upon the conventions governing the determination of references of complex signs from the references of their principal constituents. Indeed, it would be easy to give a formal definition of 'sameness of sense' for complete and incomplete signs by induction on their structure. There is nothing subjective in any of this, nothing in any way dependent on psychology. Discourse can be objective without being directly about objects.

2(*iii*) LOGICAL OBJECTS: TRUTH VALUES AND VALUE-RANGES

There is nothing problematic about the category of objects *per se*. In this respect it stands in sharp contrast to the category of functions. Objects are analogous to Aristotle's primary substances. At least, they play a roughly analogous role in Frege's ontology. Of course we must not push this analogy too far. There is a sense, for example, in which the Aristotelian category of discrete quantity is assimilated to the category of objects in Frege's system. Indeed, Frege's system is much simpler and more flexible than Aristotle's. This is partly a consequence of Frege's much narrower interests: many of the fundamental doctrines of Aristotle's metaphysics are addressed to the solution of problems which Frege does not even consider. But the simplicity of Frege's system is also partly a consequence of his very much simpler logical grammar.

For his purposes, it was not necessary for Frege either to carry out, or to preclude, a more elaborate subclassification within the category of objects. In this respect his system remains more an outline than a fully worked out theory. This is, of course, an advantage, since it makes his theory ontologically neutral, open to development in several mutually incompatible directions. And surely this is exactly what one would want in a theory of the foundations of mathematics. For such a theory ought to be above the battles fought among the various squabbling metaphysical schools, just as mathematics itself is.

But the very commodiousness of Frege's category of objects, which is the source of the flexibility and general utility of his system, is also the occasion of a serious difficulty which besets it. For it permits him to include in this category certain peculiar entities, which we might call *logical objects*, whose presence there is a potential source of embarrassment and perplexity. It is to these logical objects that I shall now turn my attention.

There are two[1] kinds of logical objects in Frege's ontology: truth values and value-ranges (*Werthverläufe*). By treating truth values as objects he simplifies his ontology and his semantics. He needs value-ranges as objects in order to carry out arithmetical definitions and proofs. The doctrine that truth values are objects could be abandoned at the price of additional complexity in his system. The notion of value-range, however, or something like it, is essential to his primary purpose of providing a foundation for mathematics.

The central difficulty here is, of course, the difficulty of answering the simple question: what sorts of things *are* these logical objects? Frege would have rejected the suggestion, put forward by Church ([1956], p. 26), for example, that we are at liberty simply to 'postulate' or 'invent' them:

> Even the mathematician cannot create things at will, any more than the geographer can; he can only discover what is there and give it a name. ([1884], p. 108)

But how can one be sure that these things *are* 'there'? Frege himself had doubts about his value-ranges even before it was pointed out to him that their inclusion in his system had led to a contradiction. And there is considerable internal evidence in the *Grundgesetze* that he never formed a clear idea of the inner or essential nature of these entities.

In the *Grundgesetze* it is merely asserted, not argued, that truth values are objects. To find Frege's argument, such as it is, for this doctrine we must turn to the essay '*Über Sinn und Bedeutung*'. His argument there is really one from analogy: if we replace the constituents of a complex name

---

[1] I have already rejected senses as logical objects.

(referring phrase) with other constituents having the same references, but possibly different senses, the reference of the complex name remains the same, although its sense may be altered. Now if we replace the constituents of a *sentence* with constituents having the same references, but possibly different senses, we obtain a sentence which though, in general, expressing a different thought will nevertheless be equivalent to the original sentence in respect of its truth or falsehood: its truth value is the same.[1] The 'thought' expressed by a sentence is thus determined by the senses of its constituent names, and therefore belongs to the realm of sense. Its truth or falsehood, however, is determined by the references of those names, and, indeed, by the truth or falsehood of its constituent sentences, when it is a compound sentence, rather than by the thoughts which they express. Truth value thus belongs to the realm of reference. May we not, then, go one step further and infer the existence of the truth values as objects which serve as the references of sentences in exactly the same way that ordinary objects serve as the references of ordinary names. Frege concluded that we ought, indeed, to do just this:

> We are driven, therefore, into accepting the ***truth value*** of a sentence as constituting its reference . . . Every declarative sentence concerned with the reference of its words is therefore to be regarded as a proper name, and its reference, if it has one, is either the True or the False. ([1892], p. 63)

He goes on, most significantly, to say

> . . . in every judgement, no matter how trivial, the step from the level of thoughts to the level of reference (the objective) has already been taken. ([1892*a*], p. 64)

It would seem, then, that the *objectivity* of a judgment consists in its being the apprehension of an *object*, namely, a truth value!

Here we can see one important reason why Frege was led to adopt the doctrine that truth values are objects. In doing so he stakes out his position in clear opposition to the 'psychologising' logicians of his day. Just as when he employs a name he intends something objective, an object, so when he employs a sentence he, quite literally, intends a genuine object, namely, a truth value. For Frege objectivity means objects.

There is a second powerful motive impelling him to adopt this doctrine. For doing so allows him to effect a considerable economy both in his ontology and in his semantics: in his ontology because he can dispense with a category of concepts, making do with functions; in his semantics because he can dispense with a separate category of sentences by assimilating them to proper names. But these savings are not accomplished without cost. For he burdens his ontology with objects, the truth values, about which he knows nothing, and can say nothing, but that they are the references of certain kinds of expressions.

---

[1] There are, of course, apparent exceptions to this in natural language, but they are not relevant to the semantics of the conceptual notation.

A similar difficulty besets his notion of value-range. But the case of value-ranges is different from that of truth values. Here Frege is in the grip of a more immediately practical concern. For what is at stake is the very possibility of carrying out his primary intention of deriving the laws and definitions of arithmetic in his system.

Let us turn to Frege's explanation of the notion of value-range in the *Grundgesetze*. What we find there is rather surprising. For he follows essentially the same procedure to characterise the notion of the value-range of a function that he employed in the *Grundlagen* to give a preliminary characterisation of the notion of the direction of a line. But the former notion is much more fundamental, and of much greater logical depth. And this leads to a crucial difference in their treatment.

He begins, in both-cases, by 'fixing the sense of an identity' in which names for the new kinds of objects—value-ranges and directions—occur on either side of the identity sign. Thus in the *Grundgesetze*, he stipulates that the expression:

'the function $\phi(\xi)$ has the same value-range as the function $\psi(\xi)$'

is to have the same reference as the expression

'the functions $\phi(\xi)$ and $\psi(\xi)$ always have the same value for the same argument'

In the conceptual notation this informal stipulation becomes Basic Law V:

$$\vdash (\grave{\epsilon}f(\epsilon) = \grave{\alpha}g(\alpha)) = (-\!\!\overset{\mathfrak{a}}{\smile}\!\!-f(\mathfrak{a}) = g(\mathfrak{a}))$$

$$(\textit{i.e.}\ \vdash (\grave{\epsilon}f(\epsilon) = \grave{\alpha}g(\alpha)) \equiv ((\forall x)\,(f(x) = g(x))), \text{ in modern notation})$$

And these arrangements are analogous to those in the *Grundlagen* where it is laid down that the expression

'the direction of the line *a* is identical to the direction of the line *b*'

has the same truth value as the expression

'line *a* is parallel to line *b*'

Now what is interesting and significant about this analogy is that it is not complete. In the *Grundlagen*, after having tentatively laid down this stipulation as characterizing the notion of direction, Frege goes on to explain why it is inadequate: it only allows us to determine the truth or falsity of an identity involving two expressions of the form 'the direction of the line —'. It does not allow us to decide, for example, whether Julius Caesar, to pick an object at random, is the direction of a particular line. In the *Grundlagen* Frege shows us how to get around this difficulty. For he replaces the above attempt at characterizing direction through 'fixing the sense of an identity' by defining the direction of a line, *a*, to be the extension of the concept *parallel to a*. In modern terms, a direction is an equivalence class of the equivalence relation *ξ is parallel to ζ*. And if we know, as presumably we do, that Julius Caesar is not the extension of any concept, then we know that he is not the direction of any line.

But do we know that Julius Caesar is not the extension of any concept? Certainly not on the basis of Basic Law V. Clearly we are facing the same difficulty here that arose with directions. But the escape route of the *Grundlagen* is no longer available. For in this case it is the notion of the extension of a concept which is itself under consideration, since, for Frege, a concept is a function, and its extension is just its value-range.

Section §10 of the *Grundgesetze* is entitled 'The value-range of a function more exactly specified'; and it is in this section that *Frege* tries to meet the difficulty just mentioned. This section constitutes one of the most opaque and puzzling passages in the book. And the root cause of this is that Frege obviously is not clear in his own mind just what sort of object a value-range is supposed to be.

He argues, first of all, that it is sufficient to determine the values which the small list of primitive functions of the *Grundgesetze* take for value-ranges on arguments. This is an extremely dubious proposition for a start. Frege employs a general notion of first level function—in quantifying over first level functions, for example—and this means that the value of any such function at such an argument must, in principle, be determined. He cannot, therefore, avoid the kind of question he raised in the *Grundlagen*: is Julius Caesar a value-range?

But we may put this difficulty aside for the moment. For it is by no means the most puzzling feature of Frege's account. He tells us that

> it is always possible to stipulate that an arbitrary value-range is to be the True and another the False. ([1893*a*], p. 48)

But why is this always possible? Frege's answer is most revealing: because we can make such a stipulation without violating Basic Law V. Surely this cannot be right. We are certainly not at liberty simply to identify different objects. We cannot, by mere fiat, make Gaius Julius Caesar identical with Gnaeus Pompeius Magnus, or the Eiffel Tower with Nelson's Column. Why, then, can we do this with truth values and value-ranges? The answer which must surely suggest itself is that we can play fast and loose with these things because we haven't any idea what they really are. The only thing we know about value-ranges, the only thing that we need require of them, is that they satisfy Basic Law V. The only thing we know about truth values is that they are the references of sentences. Neither, in short, are well-defined objects in their own right.

One must not make the mistake of underestimating the importance of this difficulty, of taking its solution to be a simple matter of 'postulating' a few entities and establishing a few conventions. To see the issue in this way would be to miss the entire thrust of Frege's thought, to overlook what was, for him, the most fundamental feature of his theory: its objectivity. For Frege, these logical objects are not mere posits or fictions; they are real entities, existing objectively, independent of human thought. He takes rival thinkers—Schröder and Dedekind, for example—severely to task for 'venturing to invent' mathematical objects. Where, he asks, are the limits of this 'inventive caprice' to be found? Where indeed.

Frege stipulates that the value range

$$\acute{\epsilon}(-\,\epsilon)$$

is to be the True and

$$\acute{\epsilon}(\epsilon = \neg\overset{\mathfrak{a}}{\smile}\!- \mathfrak{a} = \mathfrak{a})$$

the False. But it is by no means clear whether he intends these identifications to answer the question 'what are the True and the False?' or the question 'what are the value ranges $\acute{\epsilon}(-\,\epsilon)$ and $\acute{\epsilon}(\epsilon = \neg\overset{\mathfrak{a}}{\smile}\!- \mathfrak{a} = \mathfrak{a})$?'. Indeed, it appears as though he were trying to answer both questions at once.

But this clearly won't do. It is perfectly permissible to establish an artificial convention by which an hypostatised 'entity' is identified with a well-defined entity which is already known to exist. To be sure we might question the propriety of doing so in the present case because of the fundamental nature of the entities involved. But, in any case, what we have here is not the conventional identification of an invented entity with a known one, but rather the mutual identification of two invented ones. We find ourselves in the position of the man in Schopenhauer's fable who

> . . . without knowing how, is brought into a company quite unknown to him, each member of which in turn presents to him another as his friend and cousin and thus makes them sufficiently acquainted. The man himself, however, while assuring each person introduced of his pleasure at meeting him, always has on his lips the question: 'But how the deuce do I stand to the whole company?' ([1958], p. 98)

We are unable to grasp the meaning of Frege's conventions because we can form no clear idea of either of the terms being identified. And there is nothing in his exposition to suggest that Frege himself was any better off in this than we are.

All this seems the more damaging when we consider that it was technically feasible for Frege simply to have stipulated that no value-range is a truth value, that, indeed, value-ranges and truth values alike are *sui generis*, and must be distinguished, not only from one another, but from all other sorts of objects. Had he done so, and then retreated into a dignified and circumspect silence concerning the matter, he would have been more successful in disguising the fact that he really had no idea what these logical objects are in themselves.

It is particularly important that I call attention here to Frege's having explicitly rejected the conventional view that extensions of concepts are just classes, or collections of things taken in extension. In fact, Frege held the notion of a class to be a radically defective one. He took it to be a confused derivative of the more fundamental notion of a concept, and attacked Dedekind and Schröder for making it basic to their systems. But for technical reasons he could not make do with concepts alone, and thus required an object correlate of each concept, a true object which could go proxy for the concept in certain contexts.[1] These object correlates of concepts in Frege's system were their value-ranges. Unlike classes or sets, however, their essential natures remain veiled in mystery.

Frege, by his own admission, was never entirely satisfied with his doctrine of value-ranges. In Appendix II of Volume II of the *Grundgesetze* where he attempts to deal with Russell's paradox, he confesses that

> I have never attempted to conceal from myself [Basic Law V's] lack of the self-evidence which the [other laws] possess, and which must properly be demanded of a law of logic, and, in fact I pointed out this weakness in the Introduction to the first volume. ([1893*a*], p. 127)

However, as he goes on to point out, value-ranges, or something like them, are essential to his main purpose of establishing a firm foundation for arithmetic:

> I should have gladly relinquished this foundation [*i.e.* Basic Law V] if I had known of any substitute for it. And even now I do not see how Arithmetic can be scientifically founded, how numbers can be conceived as logical objects and brought under study, unless we are allowed—at least conditionally—the transition from a concept to its extension ([1893*a*], p. 127)

Value ranges, it would appear, are necessary. Unfortunately they are also impossible.

Of course we should not be surprised to discover that a notion which had so tenuous a claim to reality should lead to contradictions. Still it is a terrible irony that Frege himself should have been so dramatically confronted with the disastrous consequences of gratuitously inventing mathematical 'entities', a practice which he had so often condemned in the writings of others.

---

[1] These objects, the value-ranges, are not, however to be confused with the 'objects' which serve as references for expressions like 'the concept *horse*' or 'the function $f(x)$'. The latter are even more mysterious.

The problem confronting anyone wishing to rehabilitate Frege's system is to find an acceptable substitute for the discredited notion of value-range. And it will not do merely to postulate, as Frege did, hypothetical entities whose essence consists solely in their satisfying certain formal laws. If we wish to advance Frege's programme, we must adhere, even more closely than Frege did himself, to his ideal of objectivity. We must not posit a thing unless we have a clear notion of its essence, of what it is that makes it the sort of thing it is, and distinguishes it from all other entities. We must deal, not with artificially created fictions, but with real things whose real, objective existence can be plausibly maintained.

I shall argue in section 3 that in the notion of *set* considered as *a definite number of definite things* we have just such an acceptable substitute for Fregean value-ranges. Moreover, as I have already hinted, by incorporating this notion in the proposed new *Begriffsschrift*, we can arrive at a clearer formulation of the fundamental doctrine of Frege's ontology: the distinction between functions and objects.

2(*iv*) THE SCOPE OF THE CONCEPTUAL NOTATION

Finally, I want to call attention to a doctrine, or, rather, a feature, of Frege's system that is not often remarked upon. I refer to its ***universality***, to the fact that, in principle, everything whatsoever (and perhaps rather more besides) falls within the compass of the theory. His system is what Beth has called a *logica magna*.

This gives the theory an archaic look. For Frege's approach to logic is quite different in spirit from the modern one. Nowadays logic is seen as a general formal theory applying, not to a single universal language, but to a multiplicity of distinct but related languages: classical predicate calculi of the first or higher order. The expressions of such a language acquire meanings only when the language is provided with an interpretation or realisation of the appropriate sort. Such a realisation typically consists of a universe, or domain, of discourse for the bound and free (individual) variables of the language, together with interpretations, relative to that domain, of its individual, function, and predicate constants. But the expressions of Frege's conceptual notation all have fixed meanings relative to a single intended domain of discourse consisting of the totality of all objects whatsoever. In fact it is misleading to speak of a 'domain of discourse' here. For Frege did not regard objects as constituting a vast domain or class, except, perhaps derivatively insofar as the value-range $\acute{\epsilon}(\epsilon = \epsilon)$ corresponds to such a class. But the general notion of object is logically prior to this particular object. Moreover the gothic letters he employed in quantification are not 'variables' in the modern sense, varying over a 'domain of discourse', but merely devices for singling out the first level functions which are to occur as the arguments for the second level function $-\!\overset{\mathfrak{a}}{\smile}\!-\, \phi(\mathfrak{a})$ which corresponds to universal quantification in his system.

Modern logic is about classical predicate languages and their interpretations. As such it is better adapted to the so-called axiomatic method than is Frege's system. The modern approach permits us to define types of mathematical structures axiomatically by specifying a set of sentences—axioms—in the appropriate formal language: the realisations in which those sentences all hold true are taken to be the structures of the type being defined. In this way mathematics becomes the study of types of axiomatically defined structures: groups, rings, vector spaces, topological spaces, categories, *etc.* And corresponding to each type of mathematical structure there is the formal language in which the axioms characterizing that type of structure are expressed. We thus have one language for number theory, one for real analysis, one for the theory of groups, and so on.

Now this piecemeal approach, which is characteristic of modern logic, corresponds to something which is absolutely fundamental to the modern 'classical' standpoint in mathematics: the conviction that mathematics is essentially local in character. This view goes back to Cantor, and is something I shall discuss at length in section 3. In this sense, then, the modern system of logic is more appropriate, psychologically, at any rate, than a system like Frege's which is universal in intent. This is a great strength in practical matters. But it is a weakness when what is at issue is the foundation of mathematics as a whole. There Frege's approach is decidedly superior.

The modern approach to logic, resting, as it does, upon the general theory of models, depends, ultimately, on general set theory, in terms of which alone that general model theory can be expounded. For ordinary purposes, the set theory required can be employed 'naively' as in other branches of mathematics. But when the logical foundations of mathematics are under consideration this clearly will not do. Nor, obviously, is it possible to give an axiomatic account of set theory itself: that would lead to a vicious circle. Hence a general, universal theory of sets analogous to Frege's system is required, a system whose expressions do not require interpretation with respect to various models, but have fixed and determinate meanings. The new *Begriffsschrift* which I shall present is just such a system.[1]

3 THE NOTION OF SET

In 1885 Cantor published a brief and unfavourable review of Frege's *Grundlagen* ([1932], p. 440). This review is generally regarded as an unfair one: Dummett, for example, calls it 'savage and quite uncomprehending'. Certainly there is no evidence to be found in it that Cantor had acquired any detailed knowledge of Frege's arguments or doctrines. Indeed, it is manifestly clear that he did not give the book the careful attention it deserved. And yet, there is one short passage where Cantor's criticism seems to be very much to the point:

> The author hits upon the unhappy idea of taking what the scholastics called the 'extension of a concept' as the foundation of the concept of number; he entirely overlooks the fact that the 'extension of a concept' is, in general, something altogether undetermined quantitatively; only in certain cases is the 'extension of a concept' quantitatively determined; then, to be sure, it has a definite number, if it is finite, or a definite power, if it is infinite. However, for such a quantitative determination of the extension of a concept, the concepts 'number' and 'power' must already be given from somewhere else, and it is a reversal of the proper logical order if one undertakes to base the latter concepts on the concept 'extension of a concept' ([1932], p. 440)

Of course Cantor has got it wrong here. In the *Grundlagen* Frege does not take the 'extension of a concept' as the foundation of the concept of number, at least not in the sense Cantor seems to suppose. Indeed, Frege even says

> I attach no decisive importance to bringing in the extensions of concepts at all. ([1884], p. 117)

To be sure, in the *Grundgesetze* extensions of concepts come to play a crucial role. But Frege may perhaps be forgiven for concluding that Cantor's criticism was ill-informed and irrelevant. Nevertheless, that conclusion was profoundly mistaken. For the idea which lies behind Cantor's laconic reference to the 'quantitative determination' of the extensions of concepts is a profoundly important one. And had Frege grasped it, he might have avoided the disaster which later overtook him.

---

[1] The conventional view that set-theory should be formalised as a standard first or second order theory (*e.g.* in the language based on $\in$ and $=$) leads inevitably to the illusion that set theory is just another local theory, like arithmetic and real analysis, for example. I have called attention to the serious foundational difficulties which this gives rise to in my article [1977]. The new *Begriffsschrift* is designed as a genuinely global theory, and is therefore not subject to these difficulties. I shall return to this point in section 3(*iii*).

What does Cantor mean, then, by 'quantitative determination'? Some light is shed on this question by a passage in a letter to A. Eulenberg written near the time of the Frege review (28 February 1886):

> We are obliged to make a fundamental distinction between (*a*) The increasable actual infinite, or ***transfinitum*** (*b*) The unincreasable actual infinite or ***absolutum.***

Having made this distinction, he goes on to explain it. The 'increasability' of the transfinite he illustrates by means of examples:

> The smallest infinite ordinal number, which I designate by ω, [belongs to class ***a***, the transfinite]; for it can be enlarged or increased to the next larger ordinal number $\omega+1$, this further to $\omega+2$, and so on. Moreover, the smallest infinite power or cardinal is also a ***transfinitum***, and the same goes for the next larger cardinal number, and so on.

The absolute, however, is manifested in the absolutely infinite sequence of *all* ordinal numbers, and in the parallel sequence of *all* cardinal numbers, the $\aleph_\alpha$'s. As Cantor puts it:

> The transfinite, with its abundance of forms and configuration, points the way, of necessity, to an ***absolute***, to the 'geniune infinite' on whose magnitude there can be no increase or decrease of any sort, and which is therefore to be regarded quantitatively as an absolute maximum. The latter exceeds the capacity of human understanding, and, in particular, eludes mathematical determination. ([1932], p. 405)

The sequence of all ordinal numbers must be 'unincreasable' for otherwise it would not be the sequence of *all* ordinals. But it must therefore be 'mathematically' or 'quantitatively' indeterminable. For it is itself a well-ordered sequence, and if it could be assigned a determinate ordinal number, that number would have to lie beyond the sequence, thus 'increasing' it.

It is here, then, in his theory of the two modes of the infinite—the ***transfinite*** and the ***absolute***—that the explanation for Cantor's criticism of Frege's views is to be found. Frege's mistake consists in his failure to distinguish these modes, and in his consequent failure to see that not every concept has a 'quantitatively determined' extension.

Now we can put Cantor's objection to the doctrine of the *Grundlagen* in a form which is appropriate to what Frege actually says there: it is not possible to ascribe a (cardinal) number (*Anzahl*) to every concept, for not every concept is quantitatively determined. The concept of cardinal number itself, assuming cardinal numbers to be objects, is an example of such a quantitatively indeterminate concept.

But does this 'quantitative determination' of concepts, or of their extensions, presuppose a logically prior notion of number? In the passage I have quoted from his review Cantor clearly suggests that it does. But what did he conceive these numbers to be? He later spelled out his view on the subject in his famous monograph on transfinite numbers. And these views come as rather a disappointment. For Cantor, as for Euclid, a number is a plurality, or set, of undifferentiated 'units'. More precisely, a cardinal number is a plurality of such units, an ordinal number is a plurality of units 'given' in some definite order. Here is what he has to say on the subject of cardinal numbers:

> We will call by the name of the 'power' or 'cardinal number' of M the general concept which, by means of our active faculty of thought, arises from the aggregate M when we make abstraction from the nature of its various elements and from the order in which they are given.
>
> We denote the result of this double act of abstraction, the cardinal number or power of M, by
>
> $$\overline{\overline{M}}$$
>
> Since every single element m, if we abstract from its nature becomes a 'unit', the cardinal number $\overline{\overline{M}}$ is a definite aggregate composed of units, and has existence in our mind as an intellectual image or projection of the given aggregate M. ([1895], p. 86)

Thus cardinal numbers. The ordinal number of an aggregate M, its *order type*, is obtained, according to Cantor, by abstracting only from the natures of its elements, while retaining the order in which they are given. The order type of M is designated by '$\overline{M}$', the single bar indicating that it results from a single act of abstraction. Thus the very notation Cantor employs for these numbers serves constantly to remind us of his theory of their origin in abstraction.

Is it fanciful to see in Cantor's use of the expression 'active faculty of thought' (*aktive Denkvermögen*) an allusion to Aristotle's 'active intellect', to the *intellectus agens* of the scholastics? Perhaps. But Cantor once described himself as an 'Aristolelian realist' ([1932], p. 382). And his account of the cardinal number of a set certainly does have much in common with Aristotle's account of 'mathematical' number (in *Metaphysics* M, 1078*a* 20–32, for example), though the realistic features are more pronounced in Cantor's account than in Aristotle's, which is more ambiguous concerning the ontological status of the abstracted 'units'.

In any case, Cantor's theory of numbers as sets of abstract 'units' existing in the mind is quite untenable. For it is open to the devastating criticism which Frege directs against such views in those sections of the *Grundlagen* (§§28–44) which unfortunately seem to have escaped Cantor's attention when he prepared his review. Surely no one who has taken those arguments in, or who has read Frege's brilliantly witty attack on abstractionist theories of number in his review of Husserl's *Philosophie der Arithmetik* ([1894], p. 79), could ever seriously entertain a theory of number of the sort which Cantor held.

But where does this leave Cantor's criticism of Frege? There is still that matter of 'quantitative determination' to account for. Clearly Cantor's theory of number is not going to be of any help with that. Quite apart from Frege's objections to that theory, there is the obvious fact that Cantor's 'numbers' are obtained from the pluralities being numbered and hence are posterior to them. Still, the fundamental idea behind Cantor's criticism remains correct: Frege's mistake consists in the assumption that every concept has a quantitatively determinate extension. This assumption he later came to embody in his notion of value-ranges as objects. He came to grief as a result of his decision to take this artificial notion as the fundamental technical notion of his theory.

Even though Cantor's own theory is inadequate, it does, nonetheless, point the way to a solution of Frege's difficulties. For what is required is that Frege's untenable notion of value range be replaced by the notion of *set* considered as a *definite number of definite objects*, a 'quantitatively determined' plurality of well-defined, properly individuated things. However, what Cantor calls the 'quantitative determination' cannot be defined in terms of some more primitive or more fundamental notion of number: it must be taken as immediately given in the notion of set itself.

I regard this last point as absolutely essential. No doubt Cantor's idea that numbers are in some way prior to, or at least independent of, sets will evoke a sympathetic response in many mathematicians. Certainly as applied to *natural* numbers and *finite* sets such a view is very nearly universal. Nevertheless, I believe it to be profoundly mistaken, even in the finite case. Indeed, I believe it is a serious obstacle to the development of a satisfactory finitary theory of the foundations of arithmetic.

Clearly an examination of the relationship between the notion of set and the notion of number is called for. This is what I propose to undertake in section 3(*i*). And I believe that the result of this examination will help to show why the notion of set as a definite number of definite objects is not artificial in the sense that Frege's notion of value-range is.

3(*i*) THE NOTION OF SET AND THE NOTION OF NUMBER

The question before us is this: does the notion of set as a 'quantitatively determined' plurality presuppose, as Cantor suggests, some logically more primitive notion of number?

Our question is clearly a question for logic and ontology. But let us begin by inquiring which of these two notions is the prior one ***historically***. And let us confine our attention initially to the narrower question whether the notion of ***finite*** set presupposes the notion of ***natural*** number.

Surely, one would think, this is easily answered. After all, natural numbers, the numbers used in counting, have been with us for a long time; set theory, on the other hand, is, as everyone knows, ***the*** modern mathematical theory ***par excellence***. Rather surprisingly, however, the historical evidence suggests that the notion of finite set came first, that, indeed the original meaning of 'number' was 'finite set'.

This is the thesis which, with certain qualifications that I shall discuss later, Jacob Klein maintains in his remarkable book *Greek Mathematical Thought and the Origins of Algebra*, a book which, though intended primarily as a work of historical scholarship, is, in my view, of fundamental importance to contemporary studies in the foundations of mathematics.

What then is the picture that Klein paints for us? In Greek mathematics, a number is always a number *of*: a number of sheep, or of men, or of artifacts. It is a central thesis of Klein's book that the modern, abstract notion of number—the notion which allows us to regard the negative, rational, real, and complex numbers as 'extensions' or 'generalizations' of the natural numbers—is radically to be distinguished from the very concrete Greek notion of number as a countable plurality. Klein discusses this point in chapter 6 of his book:

> the ***arithmos*** [*i.e.* number] indicates in each case a ***definite number of definite things***. It proclaims that there are precisely so and so many of these things. It intends the things insofar as they are present in this number, and cannot, at least at first, be separated from the things at all. When Plato speaks of numbers (Republic 525D) which have 'visible and tangible bodies' the expression must be taken quite literally. For instance, in counting dogs, horses, and sheep, these processes yield dog-, horse-, and sheep-numbers. ([1968], p. 47)

According to Klein, this idea that numbers are concrete pluralities of things is not a sophisticated theory cooked up by mathematicians and philosophers, but actually represents the naive, popular view of the matter. But there is a difficulty here. For the Greeks, like us, could say that two numbers are 'the same' without thereby implying that the things numbered are identical. Klein anticipates this objection, however, and calls our attention to a key passage in Aristotle's *Physics* (*Δ*, 14, *222a. ff.*) which answers it:

> It is correct to say that the number of sheep and of dogs is the same, if that of the sheep and that of the dogs are equal; but it is not the same decad in each case, nor are the units of the one the same as the units of the other, any more than a

> scalene and an isosceles triangle are the same triangle, though they are the same figure, both being triangles; for things are called the same name if they do not differ as to the characteristic in virtue of which that name is borne, in this case the differentia of triangle. They are different as triangles, therefore, but not as figures, since they belong to one and the same subspecies of the genus *figure* . . . And so it is with numbers: for the two numbers do not differ in a numerical differentia; but the decads themselves are not the same, for ten-ness is predicated of different things—dogs in the one case and horses in the other. ([1934], p. 424)

This passage constitutes strong evidence for Klein's views. For the very question which Aristotle is trying to resolve here—how can two numbers which are distinct as pluralities be said to be 'the same'?—can only arise for someone who regards numbers, not as we do today, namely, as mere place holders in the natural number sequence, but as actual pluralities or sets of things. For Aristotle, 'sameness' as applied to numbers is not 'identity', as it would be for us; it is an equivalence relation among sets.

But I must take care not to overstate my case. Nowhere does Klein state that the Greek '*arithmos*' is just the modern 'finite set'. And, indeed, there are obstacles to the assimilation of the classical notion to the modern one. Perhaps the most serious of these is the fact that the Greek *arithmos* is irreducibly plural: for the Greeks a number was not one thing but many things. Aristotle places them in a separate ontological category of *discrete quantity* rather than in the category of *substance*. Moreover, *arithmoi*, as pluralities, never consist of fewer than two things. Sets, on the other hand, are singulars, unities—objects, in the Fregean sense. In Aristotlelian terms they would be primary substances. A set is a 'new' object, distinct from, though composed of, its elements; an *arithmos* simply 'is' the things which compose it as a plurality.

To me, the grammatical awkwardness in saying what a Greek number 'is' calls attention to a conceptual difficulty. How can we speak of 'a' or 'one' number unless it is a single thing? On the other hand, it is also odd to speak of the empty set or a one element set as a 'collection' or 'plurality'. Indeed, the empty set is an oddity however you look at it: can we really regard it as 'a definite plurality of definite things'?

Obviously both the classical notion of number and the modern notion of set give rise to ontological problems. Nevertheless the difficulties attending the idea that a number is not 'a' number, *i.e.* is not one thing, are, it seems to me, more serious than those which afflict the notion of set. The awkwardness entailed in calling a one element set a 'plurality' is, I believe, merely a verbal difficulty, and does not indicate an underlying ontological one. The problem of the empty set is more serious, but it could be solved by the simple expedient of eliminating the empty set from our ontology. This, however, would entail technical complications which it would be better to avoid if possible. Moreover, it seems to me nearly as puzzling to deny the existence of the empty set as to affirm it. I shall therefore follow the conservative course and retain it.

In any case, such differences as there are between the Greek notion of number and the modern notion of set fade into insignificance when either is compared with the modern notion of natural number, the notion that the natural numbers are the 'things' denoted by the numerals '0', '1', '2', and so on. The differences between the set and the *arithmos* should be seen as differences between cognate notions in different ontologies. The 'naive' ontology of the ancients can be systematized by Aristotle's theory; that of the moderns by Frege's. Each corresponds to a logical grammar: that of natural language in the former case, that of formal predicate calculi in the latter. We cannot expect an exact or complete correspondence between two such notions. The most we can expect is a similarity in general intention. And this, I maintain, we can most certainly discover.

Now in the light of the close relationship between Greek arithmetic and the theory of finite sets, it is most significant to observe that for the Greeks, and especially for Aristotle, the fundamental question for the foundations of arithmetic was not 'what is a number?' but 'what is a *mathematical* number?'.[1] A mathematical number is just a number in the conventional Greek sense, that is to say, a quantitatively determinate plurality. What makes it a *mathematical* number is what it is a number *of*, namely, pure, undifferentiated units (*monads*). And what makes it problematic is not its 'numberhood', but rather the ontological status of the 'units' of which it is composed. That there are numbers—pluralities of things—is, for Aristotle, in any event, beyond question. For him, the philosophical problem concerns the nature and ontological status of the pure, undifferentiated units of which *mathematical* numbers are composed.

I think it is not difficult to understand the theoretical and practical considerations which lay behind the postulation of these peculiar numbers. Theoretical arithmetic clearly requires an unlimited supply of distinct numbers. Moreover, as Plato says in the Republic (527, [18], p. 238), mathematical knowledge is knowledge 'of what eternally exists, not of anything that comes to be this or that at some time and ceases to be'. Plato met both of these requirements by postulating an infinite supply of pure, unchanging, and eternal units as the 'matter' out of which the numbers of theoretical arithmetic could be constructed.

Aristotle, on the other hand, rejected the idea that these units are separately existing substances. For him, a mathematical unit is just an ordinary object, a substance, viewed abstractly: not *qua* horse or *qua* man, but simply *qua* thing being counted. Aristotle's theory does yield a potentially infinite supply of substances, however, by virtue of the potentially infinite divisibility of extended bodies. There is thus no danger that we shall run out of numbers in his theory.

---

[1] Aristotle also considered the question 'what is an ideal number?', but came to the conclusion that there are no such things.

In modern set theory, we can always obtain a finite set larger than any given finite set, a, simply by adding the set, a, itself as a new element. The resulting set, $a \cup \{a\}$, now has one more member than the original one. Of course this works only if we take sets themselves to be objects. Since for Aristotle a set, or *arithmos*, is not one thing but many things, this method of 'generating' an infinite supply of numbers will not work. For the things which make up a number—its units—must be unities.

Let us pause for a moment and consider the practical implications of all of this. The nature of mathematical units, and of the numbers they compose, is a real puzzle, not just a 'philosophical' one. The notion eludes our grasp. How can we understand what it is to be something whose entire essence consists in simply being one thing. With ordinary, non 'mathematical' *arithmoi*, with finite sets, on the other hand, there is no such difficulty. Everyone knows perfectly well what a finite set is, though many have scruples about saying that such things really exist. These are philosophical, not mathematical, scruples, however. Doubts about the reality of sets are rather like doubts about the reality of physical objects. Serious philosophical issues are, no doubt, at stake. And yet they don't seem to make any practical difference.[1] The problem of 'mathematical objects' which confronts us is thus quite different from that which confronted the Greeks. Indeed, our problem is, in a sense, the very reverse of theirs: mathematical 'units', and the mathematical numbers which they go to make up, are, in the final analysis, unintelligible, even to those who confidently assert their reality; finite sets, in contrast, are perfectly intelligible even to those who deny their real existence. In the former case common sense tells us that there are no such things as mathematical units; in the latter, surely, it tells us that there *are* finite pluralities of things. Of course 'common sense' cannot be the final arbiter here. But it can determine which side of the argument bears the burden of proof.

It would seem, then, that the Greek notion of number survives, to be sure somewhat altered, in our notion of finite set. Moreover, by converting these numbers from mere 'pluralities' into objects, we have assured ourselves of an infinite supply of them, and thus no longer feel the need to postulate 'mathematical' numbers. In this way we manage to avoid what was for the Greeks the primary ontological problem in the foundations of arithmetic. But what about *our* notion of number. I mean the natural numbers here. Every one knows that they are just the 'things' denoted by '0', '1', '2', and so on. But what *are* these things? And do they really exist?

---

[1] Doubts about the existence of *infinite* sets are an altogether different matter, however. These do have a profound practical impact.

There is an extensive, and, for the most part, unilluminating philosophical literature devoted to these questions. Some boldly assert that numbers exist, and that they are what they are, and that's an end of the matter. Others identify them with certain sets (*e.g.* the finite von Neumann ordinals) or with equivalence classes of sets under the equivalence relation *is in one-to-one correspondence with.* For the intuitionists they are mental constructions; for the formalists, marks on paper or on blackboards. For Wittgenstein they are pieces in the language game of arithmetic. Carnap maintains that, although one can legitimately ask, say, whether there are numbers between twenty-five and forty which have only themselves and one as divisors, one cannot ask whether there are any numbers at all. Thus it would appear that, for Carnap, numbers exist 'in' arithmetic just as talking animals exist in fairy tales, or gods and wood nymphs in mythology: in all these cases we are dealing with fictions, only in arithmetic we take our fictions more seriously.

In the midst of all this speculation and controversy, Klein's views provide us with a refreshingly new perspective. He proposes to tell us, not what these numbers are, but when, and by whom, they were invented. The natural numbers, as we know them today, are not, after all, the work of God, as Kroneker would have us believe. According to Klein, they are the invention of the renaissance mathematicians who established the modern science of algebra. It was Vieta, in particular, who, by reinterpreting the arithmetic of Diophantus in his *Analytical Art*, effected the transition from the ancient to the modern standpoint.

Of course I cannot present Klein's whole case here. And without his arguments and his evidence, his conclusions will, no doubt, lose much of their impact. To appreciate fully the relevance of his ideas to the foundations of mathematics, one must travel the route by which they are established. For this, however, one must turn to Klein's own work. Nevertheless, it will prove useful for me to give a brief statement of those of his conclusions which bear upon the issues we have been discussing. According to Klein, the essential step in the passage from the ancient concept of *arithmos* to the modern concept of number consisted in the *generalization* of the concept of 'number' to include negative, rational, real, and, ultimately, complex numbers. In this generalization the concrete character of the classical 'number' is lost: it is no longer a 'numbered assemblage' or set, but has become something much more abstract. The modern descendents of the ***arithmoi***, the natural numbers, come to be regarded as just a particular sort of number. This view of natural numbers is manifested, for example, in the fact that nowadays we scarcely bother to distinguish, say, the natural number 2 from the real number 2. And what would be the point of doing so? For natural numbers and real numbers alike are, for us, nothing but abstractions, in any case.

But what is the foundational significance of all this? And where does it lead? Surely to the inescapable conclusion that the 'natural numbers' conceived as the 'things' which are designated by the numerals '0', '1', '2', and so on, are simply fictions, non-entities, mere abstractions, and that, consequently, there is no philosophical problem of what they are.[1] We don't have to explain them, nor do we have to explain them away. We don't have to 'create' them by 'basal acts of intuition' or any similar rigmarole. We don't even have to replace them by, say, finite von Neumann ordinals, and then wonder whether the 'real' meanings of arithmetical propositions have been distorted by 'translating' them into set theory.

But, it will be said, none of this is new. Why do we need Klein's scholarship to sanction such a view? After all, the idea that natural numbers, and, indeed, all 'mathematical objects', should be explained away as fictions is an idea which has gained a great deal of currency. But it is one thing to explain away the unintelligible, quite another to explain away the obvious. Here Klein's historical account of the development of the modern number concept pulls us up short. For it calls our attention to the fact that for the Greeks it was ***obvious*** that there are finite pluralities of things, that ***arithmoi*** exist. And surely it is obvious, in the same 'naive' sense that there are obviously tables and chairs, stones and trees, all the things, natural and artificial, that make up the world as we experience it in our ordinary, everyday lives. Arithmetic, as the Greeks understood it, and as we ought to understand it, is just the study of those finite pluralities: in modern terms, it is the theory of finite sets. And finite sets, whatever else they may be, are certainly not abstractions in the sense that 'natural numbers' are. Even those who chastely refuse to grant them the dignity of status as 'things' or 'objects' know perfectly well what finite sets are. But nobody has the faintest idea what 'natural numbers' are, not even those who obstinately maintain that they really exist.

An arithmetic based on so natural and so concrete a notion of number should not, in the ordinary course of things, give rise to ontological perplexity. Indeed, it is an altogether more straightforward affair than other sciences such as physics and chemistry, not to mention psychology and the social sciences. Its fundamental assumptions are, of course, metaphysical ones. But it seems none the worse for that. I believe that the widespread prejudice against any metaphysical theory of the foundations of mathematics arises out of an instinctive awareness that any metaphysical analysis of the conventionally accepted notion of number is going to arrive at 'entities' of an entirely unintelligible nature. And so it will. But so much the worse for that notion of number.

---

[1] The issue here is not just a philosophical one. We shall see (in section 4 (ii)) that this answer to the question 'What is a natural number?' has important implications for finitary arithmetic.

It should now be obvious why I think that Cantor was wrong in suggesting that the notion of 'quantitatively determined plurality' presupposes a prior notion of number. But perhaps even more important is the fact that in the notion of set we have, not a recently concocted abstraction, but a notion whose roots are to be found in the very origin of our mathematical culture. In taking this notion as a primitive notion of our new *Begriffsschrift*, therefore, we are not open to the charge of inventing mathematical entities, as Frege was when he attempted to base the original *Begriffsschrift* on the notion of value-range.

There remains, however, one conspicuous difficulty. For we must retain the option of incorporating infinite sets in our theory. But here it seems we must leave all historical precedent behind, that we may no longer claim the sanction of ancient usage. To be sure, Cantor has shown us how to develop an arithmetic of the non-finite. But such an arithmetic would have been a scandal to the Greeks for whom the concept 'infinite number' would have been a straight-forward contradiction in terms, an 'indefinite definite plurality'. The Greek *arithmos* was always conceived as finite.

Now there is a sense in which the concept of an infinite number *is* a contradiction in terms. But Cantor himself acknowledged this, and, indeed, made it the basis of his whole theory of transfinite numbers. For the distinction which Cantor draws between the finite and transfinite, on the one hand, and the absolutely infinite, on the other, corresponds, in his theory, to the classical distinction between finite and infinite. We need not despair of reconciling Cantor's transfinite arithmetic with the spirit, if not with the letter, of classical arithmetic. But if we are to do so, we must take a more thorough look at Cantor's theory of the infinite.

3(*ii*) CANTOR'S DOCTRINE OF INFINITY

There is an inevitable awkwardness in any discussion of the notion of infinity in modern set theory. For the word 'infinite' has two incompatible meanings. Its original, primary meaning is 'unbounded', 'unlimited', 'indefinitely large'; but it also has a technical, set-theoretical meaning in accordance with which we say that a *set* is infinite if it can be placed in one-to-one correspondence with one of its proper subsets. Now if by 'set' we mean 'quantitatively determinate plurality', then it is clear that no set can be 'unbounded' or 'unlimited', certainly not 'indefinitely large'. For the whole point of Cantor's theory is that any quantitatively determined plurality can be assigned a *definite* cardinality. Moreover, any such plurality is 'bounded' or 'limited' in size, for example by the cardinality of the set of all its subsets. It is, as Cantor would say, 'increasable'. Indeed, in modern set theory every set is infinitesimally small compared to most other sets, in the same way that every natural number is small compared with most other natural numbers. One might even say that this is the leading principle of the theory.

If we take the word 'finite' in its primary rather than in its technical sense, we may say that in Cantorian set theory *all* sets, even infinite ones, are finite. And this is no more paradoxical than to say that all numbers, even imaginary ones, are real.[1] In the case of real and imaginary numbers the apparent paradox is harmless, since the status of imaginary numbers is no longer controversial. But in the other case there is real mischief done, because the philosophical issue is still a live one. What is certainly true is that in both cases we are burdened with an entirely inappropriate terminology.

For Cantor, the distinction between the finite and the transfinite, on the one hand—taking those terms in their technical senses—and the 'absolute' or 'genuine' infinite, on the other, is fundamental to his entire theory:

> Another frequent confusion occurs concerning the two forms of the actual infinite, namely that in which the *transfinite* is confused with the *absolute*, whereas these concepts are strongly distinguished, insofar as the former, though a true infinite, is *still increasable*, whereas the latter is to be thought of essentially as *unincreasable* and therefore mathematically *indeterminable*. ([1932], p. 375)

Now the difference between the 'unincreasable' and 'mathematically indeterminable' absolutely infinite totality and the 'increasable' transfinite set must be more than a mere matter of size. Surely there is a qualitative difference here, a logical difference that can account for this 'undeterminability'. That difference, I maintain, is the difference between the notion of a *set* as a definite *number* of things, in general, heterogeneous in kind, and the notion of a *species* as a particular *kind* of thing, in general, indefinite in number.

What I shall call a 'species' comprises all the objects falling under a given concept, or sharing a given mathematical property, for example, the species of all von-Neumann ordinals, or the species of all singleton sets.

---

[1] Or, if one is a Carnapian: all numbers, even real ones, are imaginary.

A species, in this sense, is essentially tied to its defining property. And in this it differs radically from a set, whose essence consists in the fact that its elements are, individually and collectively, what they are, and are definite in number.

Sets are *extensional* in that there is nothing to the set whose members are $x, y, z, \ldots$ other than that its members *are* $x, y, z \ldots$; it exists independently of how, or whether, we choose to single it out. Species, on the other hand, are inextricably bound up with the manner in which they are described: they are *intentional* in character.

On this view the species of ordinal numbers, for example, is not just another aggregate which happens to be 'too big' to be assigned a number. When we investigate this species we are really investigating the logical properties inherent in the concept *ordinal number*. It is not the elements of this species, considered severally, that concern us, but the logical possibilities concealed, as it were, in the defining property which singles them out and binds them together. It is essential to the concept of ordinal number that any well-defined, quantitatively determinate collection of such numbers, any *set* of them, in short, determines a new ordinal number, their supremum, which is the smallest ordinal greater than any of the members of that collection. This internal, logical relation of the concept *ordinal number* to the concept *set* makes it logically impossible that the species of ordinal numbers should even have a 'size'.

No doubt after having struggled through the last three paragraphs the reader will better appreciate the poignancy of Cantor's remark that the absolute 'exceeds all human capacity for understanding'. Although I am less pessimistic than Cantor about this, I readily confess that I find myself straining to find the words to express what I have in mind. But I have not deceived myself into thinking that I have mounted a rigorous argument to *prove* that sets or species are this, that, or the other. There can be no question of rigorous proof here, dealing as we are, with notions so primitive and so fundamental that they cannot be analysed into simpler constituents. No, what I am doing is something altogether different: I am trying to convey, by means of 'hints and metaphors', as Frege would say, what I conceive the difference between sets and species to be. Whether this distinction is a valid one can only be judged, ultimately, by whether it enables us to make sense of things. The theory must be judged as a whole; it can be established only by dialectical, not by apodeietic, arguments.

In any case, I do not need to apologise for openly discussing this issue. The problem of infinity, in this primary sense, is there, in any event, even if we chose to ignore it. But Cantor's theory holds out the hope to us that we may escape, at least in part, the perplexities with which that problem confronts us. For if in our capacity as metaphysicians we accept his distinction between the two modes of the infinite, the transfinite and the absolute, then in our capacity as mathematicians we can safely ignore the infinite in its absolute mode. By taking as a fundamental principle of his theory the assumption that the species of natural numbers[1] form a set, that is to say, a finite plurality, he guarantees that *all* mathematical structures are finite in this primary sense.[2] The problem of the infinite, of the global structure of the mathematical universe, is, in this way, removed as a practical mathematical problem. I think this may help to explain why Cantor's theory, with its apparently extravagant ontology, has a greater appeal for the conservative majority of mathematicians than its seemingly more down-to-earth intuitionistic or finitistic rivals. Most mathematicians would, no doubt, prefer not to have to think about metaphysics. Cantor's theory banishes the metaphysical problems associated with the notion of infinity proper from mathematical practice; its rivals force us to confront them even when we do elementary arithmetic.

The notion of species, in particular, will thus necessarily play a more fundamental role in a non-Cantorian approach to the foundations of mathematics. Nevertheless, although I believe it is essential to retain the option of a non-Cantorian version of the proposed new *Begriffsschrift*, I shall not allow the notion of species to play a direct, formal role in its exposition. Species, in any case, are obviously neither Fregean objects nor Fregean functions, and a system which incorporated them directly would necessarily, therefore, go beyond the Fregean ontology. The question of the ontological status of species, and of the logical principles appropriate to them, is, in my opinion, a fundamentally important one. But it is not a question I shall consider here any further than I have done already. I shall employ species sparingly and only in the informal parts of my exposition.

3(*iii*) THE LOGICAL AND ONTOLOGICAL CONSEQUENCES OF CANTOR'S DOCTRINE

Ironically, it is to Cantor's great rival, Brouwer, that we must turn for the conceptual tools and vocabulary with which to investigate the practical consequences of Cantor's ideas. For Brouwer was the first, and so far, the only, mathematician to develop a detailed *mathematical* theory of the infinite. But before we can employ Brouwer's doctrines in the present context, we must first see the connection between his ideas and Cantor's in the proper light.

---

[1] Or more precisely, the species of finite von Neumann ordinals.

[2] For all such structures, in particular the real numbers, can be constructed from the set of natural numbers by means of standard set-theoretical techniques.

As I have already demonstrated, Cantor was fully aware of the mathematical and logical difficulties associated with the infinite, or the 'absolute' as he called it. But his theory banishes those difficulties from the centre of mathematics to its periphery. For him they constitute a problem for metaphysics, or even theology, but not for mathematics itself. Cantor's theory is really a form of finitism—to be sure, an extended and radically new form of finitism: in my article 'On the Consistency Problem for Set Theory' [1977] I christened it 'Cantorian finitism' to distinguish it from the stricter 'Hilbertian' variety. Cantorian finitism is the prevailing orthodoxy. Its peculiar strength *vis-à-vis* rival foundational doctrines is that every mathematician adopts it, in all but name, simply by mastering and employing the standard techniques of his trade, the techniques of definition and proof which are now accepted almost without question by the vast majority of mathematicians.

The central tenet of Cantorian finitism is the assumption that the species of natural numbers is 'quantitatively determined'—finite, in the primary sense—and therefore forms a set. Brouwer, as is well-known, rejected this assumption. And in this respect, Cantor's theory goes beyond Brouwer's. But in a more important sense, Brouwer's theory goes beyond Cantor's. For Brouwer's theory is a direct, mathematical theory of the infinite, whereas according to Cantor, the infinite in this 'genuine' or 'absolute' sense is 'mathematically indeterminable'. Thus Brouwer's views are at once both more conservative and more radical than Cantor's. And it is, of course, Brouwer the radical who will have the most to say to us in our present circumstances. In fact, I am convinced that Brouwer's ideas are of central importance for the foundations of our new *Begriffsschrift*, both in its strictly finitary and in its Cantorian finitary version. These ideas will illuminate two of the most important problems confronting us: the problem of the relation of our new *Begriffsschrift* to classical predicate logic, and the problem of the ontological basis for the function—object distinction.

Let us consider the logical problem first, Brouwer maintains—and this is perhaps his single most important doctrine—that classical logic is the logic of the finite. Now we have already observed that the species of Fregean objects, which includes the species of all sets, cannot be regarded as finite, even in the extended, Cantorian sense. It is, as Cantor says 'absolutely' infinite. If we accept Brouwer's dictum, then we are forced to conclude that the species of all sets, and therefore the species of all objects, cannot legitimately be regarded as a classical universe of discourse, as a domain over which the bound variable of a classical quantifier can be permitted to range.

A proposition requiring quantification over a species whose elements do not form a set (*e.g.* the species of all ordinal numbers) is really not a proposition about the elements of that species taken collectively; indeed, they *cannot* be taken collectively, because they are indefinite in number. It is a proposition about the inner logical possibilities hidden, as it were, within that defining property. It cannot just 'happen' that all the elements of such a species satisfy some property other than their defining property. There must be some inner, essential relation between that property and the defining property of the species which only logic can disclose. Of course, none of this is formally precise. But then formal precision, is out of the question here. In any case, I believe it was considerations very much like these which led Brouwer to the view that the logic of infinite species has to be radically different from classical logic.

But there is another, and I believe equally compelling reason for prohibiting classical quantification over the species of all objects. I have already remarked how, for Frege, an essential feature of his system was its universality. He believed that the category of objects must incorporate all mathematical objects and structures. And I intend that the new *Begriffsschrift* should follow the example of the old in this respect. It follows that we must include among the objects of our system *all* those structures which serve as interpretations of classical predicate languages. In particular, the domains of quantification of these interpretations must fall within the compass of our theory. Thus if our new *Begriffsschrift* is to be universal in the sense that Frege intended we must adopt the principle that *whatever can be the domain of variation of a classically quantified variable must be a set, and hence a Fregean object falling within the scope of our new Begriffsschrift.* This principle entails, of course, that we must reject classical quantification over the species of all objects. Moreover, given our identification of the notions of 'sethood' and 'finitude', this principle is just another expression of Brouwer's dictum that classical logic is the logic of the finite. (Notice again how the technical meaning of 'finite' and 'infinite' in set theory complicate the task of conveying the essence of the matter.) But it is not a question of merely appealing to Brouwer's *ipse dixit* here: this principle recommends itself by its own intrinsic plausibility. For it allows us to respond to the question 'Why does classical logic apply in these circumstances?' with the reply 'Because the elements of the domain of quantification form a set and are therefore determinate—finite—in number'.

But let us now turn to the perplexing question of the ontological status of functions. Have again, I believe, Brouwer's ideas will prove invaluable in solving our difficulties. The question which we must answer, and answer convincingly, is this: why aren't functions objects? Frege's answer, as we have seen, is that, unlike objects, functions are 'incomplete' or 'unsaturated'. Now we do not need to reject Frege's explanation out of hand in order to notice that it entails certain inconveniences and gives rise to conceptual perplexities. Some of these I have already called attention to. There is, in particular, the odd fact that, at least according to Frege, an expression which putatively denotes a function cannot, even in ordinary discourse, serve as the subject of a sentence. When such expressions are employed as grammatical subjects, Frege tells us, they designate, not the function itself, but some curious object *Doppelgänger* of the function. Thus, on Frege's account, we have, in functions, a whole ontological category of entities we can't talk about.

But, such difficulties and inconveniences aside, what are we to make of Frege's metaphor of saturation? How literally are we to take the 'gaps' in his functions. Surely there is the danger here that we are hypostatising the grammatical features of function signs and endowing them with ontological significance. We need the grammatical 'gaps', of course. For no mere assemblage of signs can constitute an articulated name or sentence. But why must the 'gaps' or 'slots' into which the simple signs fit be attached to the function signs? Might we not gather all the 'unsaturatedness' required into a single second level application function and assimilate first level function to objects? In this way we should write, say, '$Ap(\phi, a)$' instead of '$\phi(a)$'.

Now if we approach this problem from Brouwer's standpoint, the first thing we notice is that the domain of definition of a Fregean function is not a classical domain of discourse. It is, on the contrary, an absolutely infinite species, the species of all objects. Surely the indefiniteness of its domain of definition must infect the function itself. How can the function be a finite entity, that is to say, an object, under such circumstances? The essential fact about objects is that they satisfy classical identity conditions. But functions cannot satisfy such conditions for the same reasons that their domains of definition cannot be subject to classical quantification. Indeed, if classical identity could be applied to functions we could then define classical quantification over the species of all objects by means of the stipulation

$$\vdash \forall x\phi(x) = [\lambda x\phi(x) = \lambda x[x = x]]$$

All the usual classical laws of quantification could then be obtained as consequences of the classical laws of identity.

Brouwer devoted a great deal of thought to this problem of functions defined on infinite domains. His solution, which finds expression in his

Principle of Continuity, provides us with a method for treating functions non-classically as infinitary entities.

To illustrate this principle let us suppose that we can identify real numbers in the closed unit interval [0, 1] with their infinite decimal expansions.[1] Thus $\frac{1}{2}$ is ·5000 . . . or, alternatively, ·4999 . . . . Now for Brouwer a function from the unit interval into itself must be given by means of a function defined on finite initial segments of decimal expansions with such finite initial segments as values. If f is such a function and $f(\cdot a_1 \cdots a_n) = \cdot b_1 \cdots b_k$ this means that the value of the associated function on the unit interval must begin with $\cdot b_1 \cdots b_k$ whenever the argument begins with $\cdot a_1 \cdots a_n$. Of course not just any function f will do: matters must be arranged so that as the decimal expansion of the argument is developed, the decimal expansion of the value is also produced bit by bit. Now it is a consequence of this mode of definition that every function from the unit interval into itself is continuous (in fact, uniformly continuous, since the unit interval is intuitionistically compact). In Fregean terms, the decimal expansions are first level functions, and the functions from the unit interval into itself are of second level, Brouwer's *Principle of Continuity* thus says: *second level functions are continuous in their first level arguments.*

But what does this mean in the context of set theory? Just this: the value of a second level function at a first level argument must be determined by the values of its argument on the elements of some set. It cannot be required that we know all the values of the argument over its entire absolutely infinite domain in order to determine the value of the given second level function at that argument. Of course I have put this in 'operationistic' terms which aren't, strictly speaking, appropriate here. And, indeed, there are serious technical problems involved in formulating the definition of continuity in a mathematically rigorous way, problems which take us beyond the intended scope of the present essay. But it should be reasonably clear what I have in mind. If not, perhaps a few examples may help to dispel any uncertainty here. The second level replacement function is obviously continuous in this sense. This is the function, $\{f(x):x \in \xi\}$, which at a function argument, $\mathrm{f}(\xi)$, and an object argument, a, (which we may take to be a set) has as value the set, $\{\mathrm{f}(x):x \in a\}$, of all values of $\mathrm{f}(\xi)$ for arguments in a. The value of replacement at $\mathrm{f}(\xi)$ and a depends only on the values of $\mathrm{f}(\xi)$ for arguments lying in a. In a similar manner we can easily verify that the comprehension function, $\{x \in \xi:\phi(x)\}$, and the two restricted quantification functions, $(\forall x \in \xi)\phi(x)$ and $(\exists x \in \xi)\phi(x)$, are continuous in their function arguments. The unrestricted quantifiers could be only 'semi-continuous' however. $\forall x\phi(x)$ would be continuous at arguments for which it were false but discontinuous at arguments for which it were true. Conversely, $\exists x\phi(x)$ would be continuous where true and discontinuous where false.

---

[1] Such an identification cannot be justified on intuitionistic grounds, but making it here will simplify my example without doing too great violence to Brouwer's ideas.

It is not difficult to see that the Principle of Continuity precludes the possibility of introducing a second level identity relation for first level functions. Suppose, for example, that $Eq_x(\phi(x), \psi(x))$ were such a relation satisfying all the usual classical laws of identity. Then by fixing one of the function arguments we obtain a second level function of one function argument, $Eq_x(\varphi(x), \psi(x))$ which is discontinuous at $\varphi(\xi)$.

Thus the status of first level functions as non-objects, and their consequent lack of a clear principle of individuation, finds a mathematically suitable formulation in Brouwer's Principle of Continuity, which can also be taken as an expression of the absolute infinity of the category of objects.[1] As we shall see, this principle bears directly on the problem of justifying the technique of definition by recursion—finite recursion in the non-Cantorian version of our theory, transfinite recursion in its Cantorian version.

(*To be continued.*)

University of Bristol

---

[1] In this sense, perhaps, Frege's characterisation of functions as 'incomplete' is appropriate: they are 'incomplete' in the sense that infinite totalities are sometimes said to be 'incomplete', or in the sense that a 'complete' entity is a finite one.

Brit. J. Phil. Sci. 31 (1980), 255–272 *Printed in Great Britain*

Conventionalism About Space and Time*

by RICHARD SWINBURNE

In recent years philosophical writing about space and time has been permeated with a distinction between matters of fact and matters of convention. This distinction was due to Riemann and Poincaré but was developed and popularised by Reichenbach and Grünbaum. Certain claims of science are said to be truly factual and others to be merely conventional. Similar kinds of distinction are of course to be found in philosophical writing about many other issues. In general philosophy of science there is the distinction between observable entities and properties, and theoretical entities and properties; and nowadays in philosophical writing about language some writers wish to distinguish sharply between such factual matters as which noises a subject uttered at a certain time and such matters of theory as what he meant by what he said. My main concern here is with space and time, but we need to be aware of the wider background.

A matter is factual if it is expressed by a factual statement. A factual statement is one which, if true, would state a fact. Something is a fact, I suppose, for these writers, if it truly or really holds, if it is an objective feature of the world. If a factual statement is true, its negation is false; and conversely. By contrast, something is a matter of convention, in the terminology of these writers, if the statement which expresses it is not factual; and so if it is as near to the truth as its apparent negation (*i.e.* the statement which by its verbal form is the negation of the original statement, and would be its negation if the original statement were factual. If a statement is not factual and so not really making a claim, it cannot properly have a negation.) Let us call such a statement a conventional statement. A conventional statement may form a useful part of science but it does not have a truth-value on its own.[1] In conjunction with some other statement it may form a statement with a truth-value—indeed the same truth-value as the conjunction of its negation with a different statement. The two provide alternative ways of representing the same facts.

---

* An earlier version of this paper was read at a meeting of the British Society for the Philosophy of Science in May 1979. I am most grateful to those who produced valuable criticisms of it on that occasion and on similar occasions.

[1] Thus Reichenbach ([1958], p. 19) claims that in deciding on the criteria for congruence, we are making 'an arbitrary decision that is neither true nor false'.

The above seems to me the central claim which the writers cited wish to make about the statements about space and time which I shall shortly list, normally by calling them 'matters of convention'. Sometimes the writers say that alternative theories on these matters are 'co-legitimate alternatives', or 'factually co-legitimate', or, with respect to different measures of spatial and temporal intervals, that they can be adopted 'with equal factual legitimacy', *i.e.* that in this sense space or time are 'alternatively metrizable'.[1] In his later writings Adolf Grünbaum has developed more specialised senses of 'conventional' and etymologically similar terms—*e.g.* he wants to say that measurements of congruence are 'convention-laden' if and only if there is no intrinsic metric to space (*i.e.* if there are no spatial atoms of finite volume). But as Grünbaum also seems to use the other phrases quoted above and similar phrases to make the points which the other writers make, we can list him as holding that the claims which I shall discuss are matters of convention in their sense.

A very simple example will illustrate the distinction which the conventionalist has in mind. 'There are four men in this room' is a factual statement: but 'there are x men in this room' is only a conventional statement. By itself the latter has no more truth to it than 'there are $2x$ men in this room'. But the former constitutes a factual statement when conjoined to 'and $x = 4$', indeed the same factual statement as the statement made when the latter is conjoined to 'and $x = 2$'. (In formal logic the conjoining would involve the variable of the second statement being governed by the quantifier of the first statement.)

I do not dispute the utility of this distinction. My concern in this paper is with the statements which the writers concerned claim to be conventional and with the criteria which they are using in order to classify them. I list six issues, central in the philosophy of space and time, which the writers concerned have claimed to be matters of convention:

(1) Whether a body, such as the Earth, is really in motion (as opposed to, *e.g.* whether it is moving relative to the Sun, which is a matter of fact).

(2) What is the metrical geometry of a region of space—*e.g.* whether the interior angles of a rectilinear triangle in physical space sum to 180°.

(3) Whether the distance between two points *A* and *B* is or is not the same distance as that between two different points *C* and *D* (when *C* and *D* do not both lie between *A* and *B*); *i.e.* whether *AB* is congruent with *CD*.

---

[1] For these phrases see (*e.g.*) Grünbaum [1970], p. 580.

(4) Whether the temporal interval between two instants of time t_1 and t_2 is the same as that between two instants t_3 and t_4 (when t_3 and t_4 do not both lie between t_1 and t_2).

(5) What is the topology of space—*e.g.* whether a certain line in space returns to its starting point or only to a point which looks like its starting point; and so whether or not space is finite.

(6) Whether two events E_2 and E^* at different places are simultaneous (when E_2 and E^* are not in fact connectible by a signal—*viz.* when you cannot send a signal from the place of E_2 at the time of E_2 to arrive at the place of E^* before or at the same time as the occurrence of E^*, or conversely).

The grounds which the conventionalists give for holding (1) . . . (6) to be matters of convention are of course very loosely, that no one can actually observe whether things are as (1) . . . (6) claim; that so long as one is prepared to make alternative hypotheses about other unobservable matters (which one can do without the danger of being shown mistaken by observations) one can maintain which position one likes on the issues; and that there is no unknown factual truth here, because the only such truths are those which are 'in principle' checkable. The conventionalist is thus a verificationist who holds that a statement is a factual statement only if it is in some way verifiable or falsifiable by observation. What I wish to do in this paper is, first, to point out that there are different verificationist theses which have different consequences from each other about which of (1) . . . (6) are matters of convention; secondly, to show that only the stronger verificationist theses prove *any* of them to be matters of convention; and thirdly to comment briefly that the stronger verificationist theses are intuitively not nearly as plausible as the weaker ones.

Although conventionalists put forward verificationist arguments for their positions, they seldom distinguish clearly between different possible verificationist theses. Out of many different verificationist theses which could be maintained, I select four typical ones in decreasing order of strength (a thesis being stronger in so far as it pronounces fewer statements to be factual, *i.e.* rules out more statements from being factual):

[*A*] A statement is factual if and only if it entails claims which it is physically possible to observe to hold.

[*B*] A statement is factual if and only if it entails claims which it is logically possible to observe to hold.

[*C*] A statement is factual if and only if it could be confirmed or disconfirmed by claims which it is physically possible to observe to hold.

[*D*] A statement is factual if and only if it could be confirmed or disconfirmed by claims which it is logically possible to observe to hold.

As in all these theses a statement's being a 'factual' statement is tied to the notion of something being 'observable', there will be a continuum of statements from clearly factual statements to clearly non-factual statements. This is because there is a whole continuum of entities and properties from the clearly observable to the clearly non-observable—according to whether 'observing', 'in a microscope', 'in an electron-microscope', 'in a cloud chamber' *etc.* are recognised as observing. However, the kind of contrast which the writers cited wish to make seems to remain whether or not 'observing' *via* these various devices counts as observing.

The different verificationist theses suggest different criteria for logical equivalence:

[*A*] Two statements are logically equivalent if and only if they entail the same claims which it is physically possible to observe to hold. (They are then logically equivalent to the conjunction of these claims.)

[*B*] Two statements are logically equivalent if and only if they entail the same claims about what it is logically possible to observe to hold. (They are then logically equivalent to the conjunction of these claims.)

[*C*] Two statements are logically equivalent if and only if they would be equally well confirmed or disconfirmed by all claims, which it is physically possible to observe to hold.

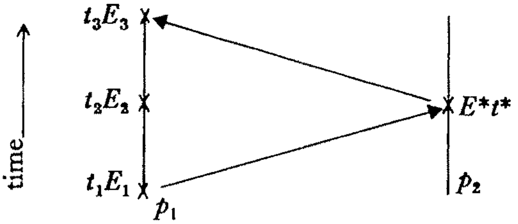
[*D*] Two statements are logically equivalent if and only if they would be equally well confirmed or disconfirmed by all claims which it is logically possible to observe to hold.

Let me bring out the differences between these different verificationist criteria. Let S entail that in circumstances C_1, F_1; and that in circumstances C_2, F_2. Let S' entail that in circumstances C_1, F_1; and that in circumstances C_2, F'_2. Suppose that these are all the claims which it is logically possible to observe to hold; that F_2 and F'_2 are incompatible; that C_1 are circumstances which it is physically possible should occur; and that C_2 are circumstances which it is not physically possible should occur. Then according to [*A*] S and S' are logically equivalent to each other and to the claim that in circumstances C_1, F_1. According to [*B*] S and S' are not logically equivalent because they make different claims about what would be observed in circumstances which it is not physically possible to realise. [*C*], like [*A*], holds that the only observational evidence relevant to determining the factual content of a statement is evidence which it is physically possible to observe to hold. But an advocate of [*C*] may hold that while two statements S and S' entail exactly the same observational consequences which it is physically possible to observe to hold—that in circumstances C_1, E_1—(and that no other evidence of observation is relevant), that evidence may nevertheless confirm one statement more than the other. This may be because one statement is simpler than the other (in postulating fewer entities, forces *etc.*, interacting in a mathematically simple way), and thus provides a simpler explanation of the evidence; and simplicity is evidence of truth. In that case [*C*] holds that S and S' are not logically equivalent. [*D*] holds that even if the only observational evidence relevant to determining the logical relations of the statements were the same, *e.g.* if F_2 and F_2' were the same, the two statements would still not be logically equivalent if there were any observable evidence (including evidence which it is not physically possible to obtain) which if it were obtained would confirm one statement rather than the other—*e.g.* on grounds of simplicity. Put loosely, [*D*] says—two statements are logically equivalent if and only if there are no conceivable circumstances in which there could be any grounds for asserting the one rather than the other.

Clearly [*A*], is a stronger criterion for logical equivalence than is [*B*], and also stronger than is [*C*], and both [*B*] and [*C*] are stronger than [*D*]; a criterion being stronger in so far as it pronounces more pairs of statements to be logically equivalent to each other.

His criterion of logical equivalence gives the conventionalist an easy method of proving a statement to be merely conventional. If one statement S is a conjunction of two statements T and Q, and a statement S', logically equivalent to S, is a conjunction of two statements T' and Q' when T is by its verbal form an apparent contrary of T' and Q of Q', then it is a matter of convention whether T or T' (Q or Q'). This is because if S is 'true', so is S', and conversely, and so then are both T and T'. If the 'truth' of T or T' were an objective fact, then it would rule out the 'truth' of the other. Hence the 'truth' of T must be such as not to rule out the 'truth' of T', *i.e.* must be a matter of convention. Put another way T and T' must be incomplete statements, not statements which can be true by themselves. In the earlier example 'there are $2x$ men in this room' is an apparent contrary of 'there are x men in this room'. But in fact both statements are incomplete, and this is revealed by the fact that when another statement is added to each, the two conjunctions are logically equivalent. The conventionalist then claims that any statement concerned with (1) . . . (6) can always be combined with some other statement so as to get a conjunction logically equivalent to another conjunction including an apparent contrary of the original statement; from which it follows that the latter must be a matter of convention.

Let us now see how this works in practice by considering how conventionalists argue for the conventionality of (6). They adduce the apparent facts, enshrined in the Special Theory of Relativity that light has a certain finite two-way velocity c (about 300,000 km/sec) relative to all inertial frames, and that no signal can travel faster than light. Now consider two points p_1 and p_2 on an inertial frame F. A light signal is sent from p_1 at time t_1 to p_2 where it arrives at t^*, and is immediately reflected back to p_1 at which it arrives at t_3. Let E_1 be the event of its emission, E^* of its reflection, and E_3 of it arriving back at p_1. Since no signal can travel faster than light, E_2 and E^* are not connectible by a signal. Hence the argument goes, the hypothesis that E^* occurs after E_2, the hypothesis that E^* is simultaneous with E_2, and the hypothesis that E^* occurs before E_2 are all compatible with the observable facts. The first hypothesis combined with the hypothesis that the one-way velocity of light from p_1 to p_2 is $< c$ (and $> c$ in the opposite direction); and the second hypothesis combined with the hypothesis that the one-way velocity in both directions is c both make the same prediction about when the signal will arrive back at p_2. And a similar point applies to the third hypothesis. Further observable consequences—*e.g.* about readings on clocks moved from p_1 to p_2—follow if we add hypotheses about how clocks are affected by transport (*e.g.* in having their rate speeded up or retarded). With different additional hypotheses added to each we can get all the same consequences which it is physically possible to observe from each of the original hypotheses. Hence, by verificationist principle [*A*], the alternative conjunctions are logically equivalent, since they hold in the same physically possible worlds. However, the component hypotheses about the temporal relations of E_2 and E^* such as that E_2 is simultaneous with E^*, are apparently contrary to one another. But they have no consequences which are physically possible to observe, by themselves. They are therefore mere matters of convention.



Not merely this; but it may plausibly be urged, the alternative conjunctions of hypotheses would be equally well confirmed or disconfirmed by all factual claims which it is physically possible to observe to hold.

It might seem that the hypothesis that E_2 is simultaneous with E^* plus the hypothesis that light has the same one-way velocity in both directions, is simpler than its rivals, for it alone of the conjunctions of hypotheses which predict the observable facts postulates that light has the same one-way velocity in both directions. It might be urged that this simpler conjunction is, qua simpler, better confirmed by the evidence which it predicts, than its apparent rivals. If we said that E^* was earlier than E_2, we would have to say that the velocity of light varied with its direction, and that is a more complicated supposition. However, all these velocities are velocities relative to F. Light has the same two-way velocity relative to all inertial frames. And if we say that E^* is earlier than E_2, then although the velocity of light from p_1 to p_2 will then be different in the two directions relative to F, there will be an inertial frame moving with some uniform velocity relative to F relative to which light will have the same one-way velocity in both directions. So, generally any hypothesis about how (within the limits set by signal-connectibility) E^* is simultaneous with, later than, or earlier than E_2 can be conjoined with a hypothesis of equal velocities in both directions relative to some inertial frame, to predict the observed evidence which I have set out. And, the argument goes on, there are no considerations of simplicity which lead us to prefer one inertial frame to any other, by which to measure velocities, for all inertial frames are on a par—for when velocities are measured relative to any inertial frame the laws of nature take their same simplest form. So each total conjunction is equally well confirmed by the stated evidence, and it is not physically possible to obtain more evidence to enable a choice to be made between them. Similar arguments would show that given the predictions of The Special Theory of Relativity, each total conjunction would be equally well confirmed or disconfirmed by other evidence. Hence by criterion [C], as well as by criterion [A], the component hypotheses about the temporal relations of E_2 and E^* are mere matters of convention. This was the kind of consideration underlying Einstein's affirmation of the 'relativity of simultaneity':

> So we see that we cannot attach any *absolute* signification to the concept of simultaneity, but that two events which, viewed from a system of coordinates, are simultaneous, can no longer be looked upon as simultaneous events when envisaged from a system which is in motion relatively to that system.[1]

And so Einstein and the textbooks of Special Relativity suggest that we should think of E_2 and E^* as 'simultaneous in F' and 'not simultaneous in F^1', and 'absolutely' neither 'simultaneous', nor 'non-simultaneous'.

---

[1] Einstein [1905], p. 42 f.

So given that there are no more relevant observable facts than those predicted by the Special Theory of Relativity, the various claims about simultaneity turn out to be matters of convention by criterion [*C*] also. However, our universe is not one in which Special Relativity holds in a pure form on the cosmological scale; and cosmology may provide good scientific grounds for preferring one inertial frame of reference to another relative to which to measure velocities (*i.e.* the laws of cosmology may take a simpler form relative to one inertial frame than relative to others). I have myself argued elsewhere[1] that there are such grounds; and that the frame by which velocities ought to be measured is the frame which has the mean motion of the galactic cluster in its neighbourhood.

Yet if we ignore this point, we can admit that criterion [*C*] as well as criterion [*A*] makes (6) a matter of convention. Clearly however, there are always facts which it is logically possible to observe which would refute any of the alternative hypotheses about the temporal relations of E_2 and E^*. For it is at best a physical truth, enshrined in Special Relativity, that no signal can travel faster than light; and it is logically possible that one might be discovered which could so travel. It might then prove possible to refute the hypothesis that E^* was earlier than E_2 by finding a signal which left p_1 at the time of E_2 and arrived at p_2 at the time of E^*. And similarly for the other rival hypotheses. So [*B*], and *a fortiori* [*D*], do not declare (5) to be a matter of convention. Why should we adopt [*A*] or [*C*]? Why should not we say instead that there is a truth there—that E_2 is simultaneous with E^*, or that it is later than E^*, or that it is earlier than E^*; and although (probably) it is not physically possible to discover what the truth is, or to have any evidence favouring one of the rival hypotheses over others, it is no matter of convention what the truth is.

Salmon claims that Reichenbach held that 'statements have the same meaning if it is *physically* impossible to get evidence to discriminate between them—*i.e.* to confirm the one and disconfirm the other'.[2] Salmon follows him in suggesting that we should judge a convention to be 'non-trivial' if it arises out of a physical rather than a logical impossibility of verification. But why should the *physical* impossibility of obtaining evidence to choose between two hypotheses show that they are not really in conflict. The grounds which the old-style logical positivist (outside the philosophy of science context) gave for tying meaning to possibility of verification concerned meaning being concerned with what we could *conceive* ourselves observing, *i.e.* the *logical* possibility of observing. The stronger criteria have the consequence that physics in showing what is physically impossible, shows that statements which looked as if they were not logically equivalent, really are. But how can a physical discovery show that sentences do or do not have some meaning or logical relation? Surely it is a linguistic, not a physical matter (not something to be discovered by connecting two wires in a laboratory), what a sentence means or what its logical relations are.

---

[1] See Swinburne [1968], chapter 11.

[2] Salmon [1969], p. 61. For Reichenbach's most developed view on this matter see Reichenbach [1953]. While indeed he here (p. 97) "advocates a definition of meaning in terms of the physical possibility of verification", he makes the odd claim that "meaning is a matter of definition" and so that alternative (but in his view less useful) definitions of meaning, *e.g.* in terms of the *logical* possibility of verification, can be given. This has the strange consequence that it is a matter of convention whether something is a matter of convention or a matter of fact.

If we do go along with the stronger criteria in this case, we have to say the kind of thing which Einstein said in the quotation given above; and if we adopt criterion [*A*] we have also to say that the factual meaning of a total conjunction of hypotheses of this kind just is the conjunction of its observable consequences, *e.g.* it is a statement about signal connecability and so forth. But intuitively statements about simultaneity are statements which may or may not be established by sending signals, but they concern something else. It seems to be very central in our thought about time that if t_2 is between t_1 and t_3 and E^* occurred between t_1 and t_3, then either E^* occurred before t_2 or it occurred after t_2 or it occured at t_2. There seems no good reason for not sticking by this way of thinking and saying that all that Special Relativity shows is that nature contains more secrets (*viz.* about simultaneity) than man can ever discover.

I turn now to (2), (3) and (4). I shall discuss only (3), but it will readily appear that exactly the same points can be made with respect to (2) and (4). (2) and (3) are closely connected. If (2) is a matter of convention, so is (3) (for a different metrical geometry entails different measurements of distance, although the converse does not hold). (4) is concerned with the claim about time analogous to that made by (3) about space. Now it looks as if not merely criterion [*A*], but criterion [*B*] as well, declare these to be matters of convention. For consider two geometrical theories of congruence (such as (3) is concerned with), one that *CD* is congruent with *AB* (*i.e.* $CD = AB$) and the other that *CD* is congruent with $2AB$. I find a wooden rod, one end of which coincides with *A* when the other coincides with *B*. I then move one end to *C*, and holding it there, make the other end coincide with *D*. Both theories remain compatible with the facts. For we can add to each a further theoretical component—to theory one, that the rod and all other measuring devices are unaffected by transport, and to theory two that the rod and all other measuring devices are doubled in size by transport. Only from such total (geometrical plus physical) theories do predictions arise. Any geometrical theory *G* about congruence between different distances (subject to the restriction in (3)) can be combined with a physical theory *P* so as to yield all the same consequences which it is logically possible to observe, as an apparently contrary geometrical theory *G′* with an apparently contrary physical theory *P′*. Hence by [*A*] and [*B*] (*G* plus *P*) is logically equivalent to (*G′* plus *P′*); and so *G* by itself and *G′* by itself are matters of convention.

The first difficulty here is that one of the (geometrical plus physical) theories may be not merely incompatible with observations but internally incoherent. For Grünbaum the meaning of 'congruent' (*i.e.* 'of the same length') is simply that of a 'spatial equality . . . predicate' determined by 'the axioms of congruence' (*i.e.* axioms which state the logical relations which hold between different statements about length)[1]; and also by the constraint that two rods which coincide at a place are congruent at that place.[2] However, it seems plausible to suppose that as they are ordinarily used the meaning of such terms is fixed not merely by their logical relations and the rules for applying them to coincident bodies but also by other standard cases of their correct application, *e.g.* that a wooden ruler moved from one side of my room to the other, which preserves congruence relations with other standard measuring rods (and so is not distorted by differential influences, *i.e.* influences such as heat which expand rods of different material to different degrees) is said to remain approximately 'of the same length'. In general, one may say, making a similar point about (4), the meaning of talk about the 'length' of temporal and spatial 'intervals' is given in part by these standard cases of correct application of such talk. And if that is so, claims that ordinary mundane objects which preserve congruence relations among themselves suddenly double in length when moved short distances seem incoherent; since they seem to violate the criteria for correct application of such terms.[3] Hence any physical theory which embodied such claims would be internally incoherent. If that is right, the fact that all standard measuring devices which coincide with *AB* coincide with *CD* would entail the fact that *AB* is approximately congruent with *CD* (and so $AB \neq 2CD$). In reply Grünbaum may claim that he is using 'congruent' in the 'modern scientific sense' in which the logical relations (plus rules for applying them to coincident bodies) alone determine meaning. But then we may reply that the initially interesting conventionalist claim (3) is no longer as interesting as it at first seemed to be. Alternatively a conventionalist may reply that his claim concerns large distances in space, where paradigm cases do not dictate correct answers. Let us so understand him, and let us see what more can be said about his claim in response to this defence.

---

[1] Grünbaum [1964], p. 27.

[2] This constraint is implicit in all his writing, although he does not spell it out explicitly in the cited passage.

[3] I argue this in my [1968], pp. 95 ff., and also in my [1970].

Criteria [*A*] and [*B*] both pronounce the two conjunctions of geometrical and physical theories logically equivalent. Whatever observable the one conjunction entails, the other entails; and this is not merely a physical but a logical truth. It is not merely physically impossible but logically impossible to check by conclusive observations whether all rods and other measuring devices moved to some distant region double in size. But criteria [*C*] and [*D*] give no clear answer as to whether the two theories are logically equivalent. It might seem quite obvious to the anti-conventionalist that the two theories differ in simplicity (the universal expansion theory being highly *ad hoc* and so complex)[1]; and that as simplicity is evidence of truth, the simpler theory is better confirmed by evidence. But even granted that in this kind of case, simplicity is evidence of truth, the argument begs the question by assuming that there are here two *different* theories which differ in simplicity. Maybe the difference of simplicity is only a difference in respect of two formulations of the same theory and if so, simplicity would not be evidence of truth. (If Heisenberg's matrix mechanics is just another way of formulating Quantum Theory to Schrödinger's wave mechanics, then even if Heisenberg's mechanics is more complex than Schrödinger's that does not show that it is less likely to be true.)

The conventionalist who uses either of criteria [*C*] and [*D*] to justify his position typically also begs the question. He admits that often when the same set of observations is predicted by two rival theories, the observations may make one theory more probable than the other. If they both predict the observations with equal accuracy, one may be simpler than the other and for that reason more likely to be true. Compatible with any finite number of data, there will be an infinite number of theories agreeing in their predictions of the data so far, but differing in their predictions for the future. Of these we judge the simpler theory and its predictions more likely to be true. This well worn point[2] was admitted in a limited way even by Reichenbach. But Reichenbach distinguished between 'inductive' and 'descriptive' simplicity. He admitted simplicity as a criterion of choice between theories which agreed in their predictions of observations made so far but differed in their predictions of subsequent observations (*e.g.* different theories about the paths on which a planet was moving, equally compatible with the positions so far observed). This kind of simplicity he called 'inductive simplicity'. But when there was no difference between two theories in the observations which they predicted, yet one was simpler (in the observable entities, properties, relations *etc.* which it postulated) the difference between them was, he said, a matter of 'descriptive simplicity'. (2), (3) and (4) are certainly on that definition cases of difference of descriptive simplicity. Then, Reichenbach claimed, inductive simplicity is, descriptive simplicity is not, evidence of truth. But why suppose that? Reichenbach did suppose that because he held that being equivalent in respect of observable facts was the same as being equivalent in respects of all facts.[1] But that is an assumption which begs all the questions; it is in fact criterion [*B*] (or perhaps [*A*]).

---

[1] I assume here that there is no background physical theory with which the hypothesis of universal expansion fits better than does its rival. If there was a relevant background physical theory, it is the simplicity of it plus any new hypothesis which is relevant, not the simplicity of a new hypothesis on its own.

[2] See (*e.g.*) Jeffreys [1973], pp. 61–4.

So, if we appeal to weaker verificationist criteria, we do not seem to be able to solve the issue of whether (2), (3) and (4) are matters of convention. Criteria [*A*] and [*B*] will solve the issue. We have seen no good grounds for adopting criterion [*A*]. What of criterion [*B*]? [*B*] has a simple doctrine—a theory is logically equivalent to the conjunction of its consequences which it is logically possible to observe.

This does not seem at first sight very plausible. (*a*) 'all swans are white' seems not to be logically equivalent to (*a*′) 'all swans so far observed and to be observed in future are white'. (*b*) 'material bodies continue to exist when unobserved', seems to be saying something contrary to (*b*′) 'material bodies exist only when observed', even if we add to each such auxiliary hypotheses that their observable consequences are the same. Likewise (*c*) 'a volume of carbon dioxide consists at all times of a large finite number of molecules, each of which consists of one atom of carbon and two atoms of oxygen', seems to be saying something contrary to (*c*′) 'However small a quantity of carbon dioxide you distinguish, it consists of qualitatively identical matter, but it changes its nature when it enters into chemical interaction or interaction with some measuring apparatus', even when appropriate auxiliary hypotheses are added.

I conclude that the stronger verificationist criteria are not at first sight very plausible, and that the weaker ones will not allow us to settle whether (2), (3) and (4) are matters of convention. (5) is in just the same situation as (2), (3) and (4). Reichenbach illustrates his claim that (5) is a matter of convention by his famous story of the spheres.[1] He imagines a world in which a man finds himself on a huge spherical surface, on which is situated his study. Having explored it all over, he finds a trapdoor in it and discovers another sphere below it, completely enclosed by the first sphere. He explores this second sphere thoroughly, and then penetrates it to find a third sphere. Eventually he comes to a fifth sphere which measurements with his rod reveal to be of the same size as the first sphere and on which he finds the geography in all respects qualitatively similar to that of the first sphere. He even finds a room qualitatively similar to the study which he left behind. To ascertain whether this room is his study, he writes a message, locks it in a drawer, passes back through the spheres to the first sphere and finds a qualitatively similar message in his study drawer. Are the two studies and so the two spheres numerically identical or merely qualitatively similar? If they are numerically identical, then space has a non-Euclidean topology. But if they are different, space may be Euclidean; and in that case there is causal action at a distance and all goings-on on sphere 1 are replicated on sphere 5 (and sphere 9, 13, and so on *ad infinitum*). According to Reichenbach, N (the claim that space has a certain non-Euclidean topology) plus P (some claim about the behaviour of bodies, including the claim that to some extent bodies preserve their appearance and that goings-on are not replicated) is logically equivalent to E (the claim that space has a Euclidean topology) plus G (some claim about the behaviour of bodies, including the claim that goings-on are replicated). Hence N by itself or E by itself are mere matters of convention. Reichenbach's story is meant to illustrate the general thesis that you can make any claim you like about the topology of physical space, so long as you are prepared also to make claims about strange physical forces complicating effects.

---

[1] See Reichenbach [1938]. He writes (p. 374), 'There are cases in which the simplicity of a theory is nothing but a matter of taste or of economy. These are cases in which the theories compared are logically equivalent, *i.e.* correspond in all observable facts'.

Now, as with congruence claims, it does seem not merely physically, but logically impossible to make observations entailed by one of such conjunctions of geometrical and physical theories and not by the other. So not merely criterion [A] but criterion [B] declares (N and P) to be logically equivalent to (E and G), and so N and E by themselves to be matters of convention. However, the hypothesis of replication looks complex, and it is natural to urge that, qua simpler, (N and P) is more likely to be true than (E and G); and so that the evidence of observation which both conjunctions entail confirms the former more than the latter, and so the criterion [C] declares them not to be logically equivalent. More generally, it may seem that there will always be physically possible observations such that if you are to save some geometrical hypothesis in the face of them, you need so to complicate your physics that it becomes implausible. It looks as if criterion [*C*], and so *a fortiori* criterion [*D*], declare (5) not to be a matter of convention. But an opponent may always urge that this is begging the question by assuming that conjunctions such as (*N* and *P*) and (*E* and *G*) are different theories; for simplicity is only a criterion of choice between theories which are not logically equivalent.

---

[1] *Op. cit.* section 12.

The implausibility of criteria [*A*] and [*B*] as criteria of logical equivalence, and the fact that criterion [*C*], and *a fortiori* criterion [*D*] are of little use, suggests that verificationist criteria are of no use for settling whether statements are logically equivalent, and so settling whether (2), (3), (4) and (5) are matters of convention. I do not know of any other general principle which will settle the issues quickly.[1] All that can be done, I suggest, is to spell out in detail how a Universe in which one conjunction of theories held would differ from a Universe in which an alternative conjunction of theories held; and thereby show that really the two Universes are the same or that really they are different. Thus, one example of a Universe in which one conjunction of theories of type (3) held would be a Universe in which the distances between galaxies at a given cosmic time increase as you get further away from the Earth but there are physical forces at work which expand rods to a similar amount, and increase all two-way signal velocities also so that this is not noticed. One can continue spelling out the nature of these forces, and how actual distance differs from measured distance. By contrast a Universe in which the other conjunction of theories holds would be a Universe in which the distances between galaxies are at a given cosmic time approximately the same, measuring rods are not subject to expanding forces *etc.* As the spelling-out goes on, it may clearly click that the Universes differ or that they are the same. That may seem feeble, but is not all argument an attempt to settle something questionable by appealing to something more obvious? This is merely a case of that. It might be thought that the logical equivalence of the two Universes (and so the conventionality of (3)) could be proved by showing that the affirmation of existence of a Universe of one kind and the denial of the existence of any Universe of the other kind generated a contradiction. And indeed so it could. But I cannot see how to prove that there is a contradiction until you have proved by some other method (*e.g.* my feeble spelling out) that there is no difference between the two Universes. For myself, I can only find that this attempt to spell out a difference between alternative theories in this field fails, and that I cannot make any sense of the supposed distinction between distance and what would be measured by actual rods.

---

[1] Quine [1975] has suggested a test of a non-verificationist kind for determining whether two theories which predict the same observations are logically equivalent. The test, roughly, is that two theories are logically equivalent if we can get the one from the other by substituting predicates throughout. Thus our present theory of physics and chemistry is logically equivalent to one obtained from that theory by replacing all occurrences in it of 'electron' by 'molecule', and all occurrences of 'molecule' by 'electron'. However, as Quine states, this test is concerned only with predicates which (p. 319) 'do not figure essentially in any observation sentences'. But theories normally contain a large number of terms which do occur in observation sentences—*e.g.* 'contains', 'is composed of', 'particle', 'wave' and in our examples 'distance', 'length', 'moves' and 'at rest'. Whether these terms occur 'essentially' raises a host of difficulties which are no easier to solve than our original difficulties. Quine has no doubt provided a sufficient condition for logical equivalence, but it is in no way obvious that he has provided a useful test, let alone that he has provided a necessary condition for logical equivalence. He admits that in practice there are great difficulties in applying his test. The real difficulty is that theories use terms with extra-theoretical meaning (as they must, if they are to be informative) and it is far from clear to what the use of these terms commits a speaker.

(5) on the other hand I find different. There does seem to me to be a clear difference between a non-Euclidean universe (N_1) and a Universe with Euclidean geometry in which all effects are replicated (E_1). I can try to make this clear to someone who does not see the difference by spelling out what is involved in the latter. I ask him to consider another Euclidean universe of nested spheres (E_2) in which effects on Sphere 1 are replicated in almost every detail in Spheres 5, 9, 13 *etc.*, but not quite. Such a universe would be entirely different in character from the non-Euclidean universe N_1 in which you reach Sphere 1 again by passing from Sphere 1 through spheres 2, 3 and 4. But the universe E_1 in which effects were perfectly replicated would only be very slightly different from E_2. So E_1 could not be the same as N_1. In this way I try to draw a contrast between Euclidean and non-Euclidean universes. Others may not accept that I have succeeded in making a contrast, but in that case all that can be done is to go on drawing out the differences and similarities until it does click that the universes are the same or different. Whether or not there are real distinctions to be made between different kinds of relevant Universe, and so whether (2), (3), (4) and (5) are matters of convention is something which can only be shown by detailed spelling-out of what the alternatives amount to—and the process may not immediately lead everyone to the right conclusion. The arguments here may be correct without being immediately persuasive.

(1) is in a different position from (2), (3), (4) and (5) in that not merely can criteria [*A*] and [*B*] be deployed to show it to be a matter of convention, but criterion [*C*] clearly can as well. For it follows from the Special Theory of Relativity, as from Newtonian mechanics, that there will be no observable phenomena which are more plausibly explained by supposing

that a body is at rest relative to an absolute rest frame than by supposing that it is in uniform motion relative to that frame. This is because Special Theory plus the theory that a body *B* is in motion relative to an absolute rest frame predicts exactly the same consequences as Special Theory plus the theory that B is at rest relative to that frame, and neither conjunction of theories is simpler than the other. Hence, even if we can talk of an absolute rest frame and so of absolute rest and motion, the two theories that *B* is in absolute motion and that *B* is at rest absolutely will be equally well confirmed or disconfirmed by all factual claims which it is physically possible to observe to hold,[1] and hence by criterion [*C*] they are the same theory.

However, it is logically possible that there be observable phenomena incompatible with Special Theory which are most plausibly explained by a rival physics *M*.[2] It could be that *M* plus the theory that *B* was in absolute uniform motion yielded different observable consequences from the consequences predicted by *M* plus the theory that *B* was at rest absolutely; that these latter consequences were observed, and that the only way to save the theory that *B* was in absolute motion was by making *M* into a more complex theory *M'*. (*M* plus the theory that *B* was absolutely at rest) might predict all the observations which it was logically possible to make, as did (*M'* plus the theory that *B* was in absolute motion). Yet because *M* was simpler than *M'*, one might claim that the observations confirmed the former conjunction more than the latter, and so that these were not logically equivalent (and so that (1) was not a matter of convention). But this argument already assumes what it seeks to prove—that the conjunctions of theories are different—for otherwise no evidence would confirm one against the other. We are back with the same difficulty as with (2), (3), (4) and (5). It remains the case that there is no knock-down proof that the rival theories are the same, even if you adopt the weakest verificationist criterion. Again the only way to settle things is to spell out in detail what it would be like for each of the conjunctions to hold, and see if they obviously describe different states of affairs. For myself I now find it difficult to see that they do—whatever the *M* or *M'*, for I cannot make sense of the suggested contrast between the alternatives '*B* is at rest absolutely' and '*B* is in uniform absolute motion', unless these are added to physical theories which give content to the 'absolute'. I cannot understand what is meant by saying that something is moving, unless it is implied or stated relative to what it is moving. Certainly there may be a substance such as the aether filling all space, and thus a difference between bodies which move relative to the aether and bodies which do not. But the moment such a substance is given as the relatum of motion, the question can be asked whether *it* is moving absolutely or not; and its absolute motion cannot be just motion relative to something in space. (For if it were, we could raise the question about the motion of that something.) And even if space itself is a substance in *some* sense, I just cannot understand the differences between a body which moves relative to it and a body which does not. But this may be simply the result of a lack of philosophical insight on my part, and maybe more detailed spelling out of the hypothesis would reveal to me the difference. I do not wish to deny that there is a difference between uniform motion relative to an inertial frame, and acceleration relative to an inertial frame. Nor that (if the laws of nature were different) there could be, for all F and F' when F' is in motion relative to F, an observable difference between the laws governing the behaviour of bodies when in motion relative to F and the laws governing their behaviour when in motion relative to F'. I am merely denying that an explanation of this difference in terms of F being at rest 'absolutely' describes a state of affairs intelligible without a physical theory about 'absolute motion' to give meaning to this otherwise empty notion.

---

[1] A similar conclusion also holds if we consider not Special Theory, but the modern cosmology of the Robertson–Walker line element. See Swinburne [1968], p. 59 f.

[2] See Swinburne [1968], p. 65 f. for a development of such a physics.

The main point of this paper was not however to solve the problems of whether (1), (2), (3), (4), (5) and (6) are matters of convention. It was rather to point out that if you attempt to solve them as conventionalists always have done by an appeal to verficationist criteria you would need some implausibly strong criteria to show any of them to be matters of convention; that weaker verificationist criteria show (6) not to be a matter of convention, but do not settle matters for (1), (2), (3), (4) and (5); and that the only way to settle the status of these is by the non-verificationist method of careful spelling-out of alternatives, which does not allow things to be settled by quick knock-down arguments.

University of Keele

REFERENCES

EINSTEIN, A. [1905]: 'On the Electrodynamics of Moving Bodies', in H. A. Lorentz *et al.* (ed.) *The Principle of Relativity*: London, Methuen and Co., 1923.

GRÜNBAUM, A. [1964]: *Philosophical Problems of Space and Time.* London: Routledge and Kegan Paul Ltd.

GRÜNBAUM, A. [1970]: 'Space, Time and Falsifiability, Part I', *Philosophy of Science,* **37,** 469–588.

JEFFREYS, H. [1973]: *Scientific Inference,* Third edition. Cambridge: Cambridge University Press.

Discussions

ZAHAR ON MACH, EINSTEIN AND MODERN SCIENCE

1 In his [1977] note (numbers in brackets will from now on mean pages of this note) Elie Zahar states

(*A*) that Einstein, when introducing special relativity, violated 'cardinal' principles of Mach's philosophy (205);
(*B*) that without the violation 'the special theory of relativity would never have seen the light of the day' (195); and he infers
(*C*) that Mach's philosophy 'was largely irrelevant to the development of modern physics' (195).

Now the fact that Einstein violated *some* 'principles' of Mach's philosophy does not show that his ideas were independent of *other* principles of that philosophy; nor does it show that the philosophy did not influence other parts of modern physics, such as the quantum theory: Zahar's arguments do not establish *C*.

2 They also fail to establish *A*.

Zahar mentions four conflicts between Mach and Einstein's ideas and procedures. They are (*A1*): relativity 'is a *causal* theory in the traditional sense of the word' (203) *i.e.* a theory involving an assymmetry while Mach's functional approach eliminates all assymmetries (202). (*A2*) Einstein's definition of simultaneity violates 'one of Mach's cardinal methodological principles' (205) and it is also 'radically different' (203) from Mach's definition of mass. (*A3*) the special theory of relativity uses inertial frames which, being unobservable, are anathema to Mach. 'It is therefore small wonder' writes Zahar on this point 'that Mach should have disowned relativity' (206). Inertial frames are also needed to make sense of general covariance hence Mach, when arguing for general covariance 'was violating a tenet of his own branch of positivism' (212). (*A4*) for Einstein (*a*) 'laws have an ontological status over and above the observables which they correlate with one another' (212) and (*b*) they must have such special status, for otherwise the arguments for general covariance again fail to make sense.

3 But Mach, far from treating inertial systems as metaphysical monsters examined their properties and watched with interest and approval the attempt to make them a basis for astronomical measurements. He regards Corollary V of Newton's *Principia*[1] 'as providing the only practically useful (and most likely approximate) inertial system[2] and he presents L. Lange's 'attractive idea' concerning the definition of such systems.[3] The definition, Mach remarks, separates

---

[1] Cajori (*ed.*) [1962], 20 f.

[2] Mach [1933].

[3] Mach [1933], p. 233. Mach mentions some technical difficulties of Lange's definition, but he nowhere regards it as metaphysical.

the 'conventional' and the empirical elements of the law of inertia which consists in a 'restriction of the kinematically possible manifold'. Mach also describes Seeliger's research concerning the relation between inertial systems and the 'empirical coordinate systems actually in use in astronomy'. The attempts to establish 'ephemeris time' and a fundamental reference system[1] which are attempts to make inertial systems the basis of astronomical observations are well known to him and are regarded by him as important empirical investigations: Mach's aversion to relativity has no connection with the occurrence of inertial systems in it.

4 *A4a* is also incorrect. In his essay 'Physics and Reality'[2] which criticises the quantum theory for its '*incomplete* representation of real things',[3] Einstein explains what he means by 'real existence':

> Out of the multitude of our sense experience we take, mentally and arbitrarily, certain repeatedly occurring complexes of sense impressions . . . and correlate to them a concept—the concept of the bodily object. Considered logically this concept is not identical with the totality of sense impressions referred to; but it is a free creation of the human (or animal) mind. On the other hand this concept owes its meaning and its justification exclusively to the totality of the sense impressions which we associate with it.
>
> The second step is to be found in the fact that, in our thinking (which determines our expectations), we attribute to this concept of a bodily object a significance which is to a high degree independent of the sense impressions which originally gave rise to it. This is what we mean when we attribute to the bodily object a 'real existence'.[4]

This passage is pure Mach[5] as can be seen from what Mach says about concepts in *Erkenntnis und Irrtum*[6]: according to Einstein the issue in quantum theory is not an 'ontological' issue (212) whatever that means—it is an issue over different modes of 'arbitrarily' ordering sense data.[7]

---

[1] *Cf.* the explanations given in the *Explanatory Supplement to the Astronomical Ephemeris and Nautical Almanac*, London, 1961.

[2] Einstein [1936] p. 221; reprinted in [1954], 290 ff. I shall quote from the latter source.

[3] p. 325 f. Italics in the original.

[4] p. 291.

[5] Except that Mach's account is more realistic than Einstein's: Mach ascribes to a natural development what Einstein regards as a 'free creation of the human (or animal) mind'. For Einstein concepts are freely at our disposal. For Mach their formation obeys psychological laws.

[6] Both steps are explained already in Mach's earlier writings, for example in his [1896], pp. 207, 213 (first presented at a meeting of the Vienna Academy in 1882).

[7] The same applies to Boltzmann's defence of atomism. 'Hertz made it very clear to physicists' Boltzmann said in 1899 (See his [1905], p. 215) 'what philosophers most likely have said long ago, that a theory cannot be an objective thing, in agreement with nature, but can at most be a mental picture of phenomena, being related to them like the sign to the thing signified'. (Hertz had used the term *Scheinbild*—phantom picture). Boltzmann, accordingly, regarded theories such as the atomic theory as 'mental pictures' (p. 142); his defence of atomism was not an 'ontological' argument, but an argument

Confronted with Machian utterances by scientists Zahar speaks (more than once) of 'paying lip service to Machian positivism' (195) and he adds that scientists often do one thing and say another (197): obviously it needs a philosopher to find out what is really going on. But the above quotation cannot be explained away in this simple fashion. It is part of a long article on the situation in physics; it prepares a criticism of the quantum theory; it does not belong to Einstein's early period where some 'lip service' to Mach *might* have occurred (though it is not at all clear why Einstein should pay 'lip service' to any philosophy) but to the 'reactionary' part of Einstein's life when to outsiders he more than ever seemed to be an 'old fashioned realist' (195). The quotation shows that he was 'old fashioned' in the sense of Mach: his objections are conceptual, not 'ontological'. And the article as a whole shows that the saying-doing dichotomy is anything but realistic in Einstein's case: it is very implausible to assume that Einstein, when directing attention to epistemological matters closely connected with the physical research he was engaged in should have failed to know what he was talking about.[1] A look at *A4b* shows that he certainly did not on the matter at issue.

5 Zahar says that arguments about covariance involve the behaviour of laws in different reference systems and therefore assume 'that laws and their form possess an objective status' (212); they also assume coordinate systems which are forbidden by Mach and run counter to the principle, allegedly defended by Mach, that the simplest formulation of a law is the best: 'there is no good reason, from a Machian viewpoint, why the choice of a certain coordinate system should not prove more convenient than another' (212). The last point is of course true and this is why *particular problems* such as the relativistic advance of the perihelion of Mercury are calculated in special reference systems (Schwarzschild's calculation). But if the purpose is to get laws connecting the special cases then 'physics, for reasons of economy, often gives us equations between functions of the primary variables and not equations between the primary variables themselves'[2] and the task is then to make *these* equations as simple as possible: Zahar has a rather simpleminded view of Mach's conception of economy. But does not

---

concerning the superiority of some pictures over others and he emphasised that arguments of this kind were perfectly in agreement with Mach's general philosophy. Mach, on the other hand ('whose relevant writings greatly contributed to the clarification of my own views'—*op.cit.*, p. 142, n. 1) had no objections to even the most outlandish ones of Maxwell's models and ascribed to them 'Maxwell's extraordinary success': Mach [1917] (first published in 1905), p. 229 f. These and other matters are explained in detail in M. S. Curd's thesis, *Ludwig Boltzmann's Philosophy of Science*, Pittsburgh, 1978, where the 'hagiographic' account of the 'realist' Boltzmann's battle against shortsighted positivists such as Mach and Ostwald is laid to rest once and for all.

[1] Einstein is the last of the late nineteenth century scientist–philosophers who like Maxwell, Hertz, Mach, Boltzmann and Duhem designed philosophies for advancing science and used (corrected) these philosophies in their research. Today philosophers are outsiders with correspondingly unrealistic views of history and scientific method. Zahar wants to convince us that these outsiders have a better view of science than the most outstanding scientists. His article tells otherwise.

[2] 'Die oekonomische Natur der Physikalischen Forschung' quoted after Mach [1896], p. 228.

the formation of functions of laws assume that laws have an 'objective status'? Yes—but not in Zahar's sense: talking about blueprints assumes that blueprints are 'objective' but not in the sense that they have an 'ontological status *over and above*' the things they are blueprints of. The 'objectivity' of a blueprint, however, can be easily explained in Einstein's (*i.e.* Mach's) manner, by giving an account of the way in which its elements hang together. Reference systems, finally, are elementary for a Machian (see the refutation of *A3 above*). This concludes the refutation of *A4*.

6 There remains *A2*. The answer is here that Zahar has interpreted a difference between definitions (definition of mass on the one side, definition of simultaneity on the other) as a difference between philosophies (Mach's and Einstein's). But Mach, far from subjecting definitions to rules 1–3 (202*f*) distinguishes various types of definitions, purely 'nominal stipulative' (207) definitions included. For example, he considers the following definitions of a time metric: (*a*) *t* is proportional to the angle of the earth with respect to some inertial system; (*b*) *t* is proportional to the logarithm of this angle, and he comments on the arbitrary and rival-hypotheses-compatible (207) nature of the stipulations.[1] Einstein's definition of simultaneity falls into this category and is therefore anything but a 'viola[tion] . . . of Mach's cardinal methodological principle' (205). Nor did Mach ever subject definitions to condition 2 (203). Quite the contrary—in his definition of mass he explicitly states the dependence of the definition on 'the existence of a special, acceleration—determining quality of bodies'[2] *i.e.* on what Zahar calls theoretical presuppositions. Condition 3 is not Machian either. For although Mach objects to what one might call direct circularity (explicit definitions which contain the thing to be defined) he is well aware of the fact that concepts, being dependent on each other in many ways, cannot always be explained in isolation. Zahar's condition 1, however, (201) can be refuted at once: 'Which concepts we form' writes Mach,[3] 'how we delimit them with respect to each other—all that can be decided only by the needs of practice, or of science': Mach is much more liberal than modern philosophers of science and his 'cardinal methodological principle' (205) is nothing but a chimaera in Zahar's head.

A1, finally, overlooks that Mach paid great attention to unidirectional processes[4] and occasionally regarded them as fundamental: not all connections are symmetrical and those which are 'are secondary, not basic'.[5] Zahar connects Mach's alleged failure to account for assymmetry to his 'conceptual—as opposed to the propositional—approach' (202). He overlooks that Mach's concepts are concepts in use,[6] *i.e.* have propositional content while the arrow of implication (203) being reducible to disjunction, contains no asymmetry. What is needed in matters of symmetry *vs.* asymmetry is not an emancipation from an alleged

---

[1] Mach [1917], p. 435. *Cf.* also Mach's comments on the conventional elements of the law of inertia mentioned above.

[2] Mach [1933], p. 211.

[3] Mach [1917], p. 129.

[4] Mach [1933], p. 218 and Mach [1922], 286 f., especially 286, lines 12 ff.

[5] Mach [1917], p. 443.

[6] 126 ff. and *passim*: Mach has a decidedly constructivist (in the sense of mathematical constructivism) attitude towards concepts.

'conceptual' philosophy and its replacement by a Popperian propositionalism,[1] what is needed is 'more research' concerning natural processes.[2]

7 There are many other mistakes. Zahar: 'Mach thought . . . he could boil everything down to spatial measurements' (206). Mach: 'there is an unchangeable physical time object just as there is an unchangeable physical space object—the rigid body'.[3] Zahar: '*Elements* . . . are the simplest component parts of sensations' (200—his emphasis); Mach: 'elements are sensations *only insofar*' as we consider their dependence on a particular complex of elements, *viz.* our body; '*they are simultaneously physical objects*, *viz.* insofar as we consider other functional dependencies'.[4] Disregarding the dependencies elements are therefore as much physical objects as they are sensations.[5] Zahar's biggest mistake which makes manifest his ignorance concerning Mach's philosophy is to describe it as 'strictly positivistic' (198).[6] Already the title of Mach's first major work should have given him pause. An author who writes an *Analysis* of sensations obviously does not take sensations as basic. And, indeed, Mach insists on an analysis of sensations including an examination of their dependence on physiological conditions,[7] he regards results of such an analysis as 'temporary'[8] and sensation talk as involving a 'one sided theory'[9]; he asserts that introspection does not suffice as an instrument of analysis but must be combined with a physiological study[10]; he points out that our ideas affect not only 'the development of *experience*' but 'enrich every . . . fact'[11]; he objects to Mill's aversion to hypotheses which 'favours what has been found over what has still to be discovered'[12]; he encourages the researcher 'not to be too fearsome when introducing hypotheses. Quite the contrary, a little boldness in these matters furthers research'[13]; he quotes with approval Priestley's observation that 'very lame and imperfect theories are sufficient to suggest useful experiments while serving to correct those theories and give birth to others more perfect. These then occasion farther experiments *which bring us still nearer to the truth*, and in this method of approximation we must be content to proceed . . .'[14]; he describes himself as a

---

[1] Zahar, like all good Popperians, ascribes even the most trite and useless idea to Popper. In the case of propositionalism the origin is clear: it is Aristotle.

[2] Mach [1933], p. 219.

[3] Mach [1917], p. 393.

[4] Mach [1922], p. 13, my emphasis.

[5] Mach seems inclined to count some of Duhem's 'qualities' among the elements. *Cf.* his introduction to the German translation of Duhem's main philosophical work, *Ziel und Struktur der Physikalischen Theorien*, Leipzig, 1908. But 'qualities', for Duhem, are physical primitives such as currents, charges and so on.

[6] Carnap [1928], p. 87, notices with surprise that Mach's 'basis does not consist of events belonging to the perceiving subject' ('die nichteigenpsychische Basis') and point out that this feature 'constitutes a disturbing element in the total picture of his (Mach's) ideas'. Obviously he has formed a false conception of this picture—and he now notices difficulties. No such difficulties exist for the Popperian Zahar.

[7] *Cf.* Mach [1896], p. 228.

[8] Mach [1922], p. 24.

[9] 18.

[10] Mach [1896], p. 228.

[11] Mach [1917], p. 245.

[12] 249.

[13] 247.

[14] 240 f., my emphasis.

'non-empiricist' or 'not-only-empiricist'[1] and, considering his reconstruction scheme of science in terms of hypothetical elements (*not* sensations) he warns that 'there is no need to turn *this passing point of view* into a system for life whose slaves we then become . . .'.[2] There is a great distance indeed between Mach's 'philosophy'[3] and Zahar's rather unbelievable distortion of it. To ascribe such views to Mach shows not only an astonishing ignorance of relevant historical material; to assume that a scientist–philosopher and social critic of Mach's calibre[4] should in the end have come up with such a sophomoric theory shows in addition an enormous lack of imagination and historical understanding.[5]

8 There is, then, *no conflict* between Mach's philosophy and the (genesis of the) special theory of relativity. But is there a positive relation? Are there elements in Einstein's early research that show a similarity to Mach's philosophy and may therefore be regarded as evidence of Machian influences? On the other hand—what are we to make of Mach's criticism of relativity? I shall now give some very tentative answers to these questions.

To explain Mach's criticism I draw attention to two passages contained in a longer quotation from Mach by Zahar (197). These passages both refute Zahar and also point to an answer rather different from Zahar's. The fact that Zahar quotes them and yet continues on his own line of attack shows to what an extent Popperian mythology is capable of blinding its followers.

Mach discredits relativity because he finds it 'growing more and more dogmatical' (first reason) and because of 'particular reasons', 'based on the physiology of the senses, the theoretical ideas and, above all, the conceptions resulting from experiments' (second reason).

I connect the second reason with Mach's attempt to find a point of view that is not restricted to a particular subject such as physics, or psychology 'but can be retained in all domains'.[1] Or, to express it differently, but still in agreement with Mach's intentions[2]: he sets himself the task to find theories that cut across traditional boundaries such as the boundary between physics and psychology without being beset by the mind-body problem. What have 'experiments' got to do with the construction of such a theory?[3] In an earlier essay Mach comments that 'the mental field', *i.e.* the domain where thoughts, emotions *etc.* appear and disappear 'can never be fully explored by introspection. But introspection combined with physiological research which explores the physical connections can put this field clearly before us and only thereby make us acquainted with our inner being':[4] introspection does not suffice; it grasps only part of 'mental events'; their whole nature is revealed by introspection plus physiology. The partially physical character of the research leads to the expectation that our 'inner being' ***may without break merge*** into the physical world thus providing a unified account of 'mental' and 'physical' events and laws. A purely physical theory of space (or spacetime) does not fit into such a research programme. This is the 'particular reason'.[5]

---

[1] Hugo Dingler [1924], p. 61, footnote. *Cf.* also notebooks, vol. ii, p. 16 from back (contained in Dingler's book) where Mach puts sensations in their place and wants to free science from their dominance: Mach criticised not only realism, but *all* one sided philosophies. Zahar refers to Dingler's book, but seems to have read only a tiny part of it.

[2] Mach [1896], p. 226. One should compare this quotation with the Einstein quotations in section 9.

[3] Mach refused to regard his hints or '*aperçus*' ([1922], p. 39) as a philosophy: 'I am a scientist, not a philosopher . . . and I think I am innocent of the fact that people have regarded me as an idealist, Berkeleyan, even a materialist.' *Loc. cit.* 'Above all there is no Machian philosophy' he writes in the introduction to his [1917] p. vii—'there is, at most, a methodology of science and a psychology of knowledge *and like all scientific theories these things must be regarded as preliminary and incomplete attempts*' (my emphasis). There are many philosophers with whom Mach feels he has 'points of contact', Spinoza among them ([1922], p. 38). Note how much more dogmatic critical rationalists sound when pronouncing on their 'philosophy'.

[4] 'I don't think anyone ever gave me so strong an impression of pure intellectual genius. He apparently has read everything and thought about everything, and has an absolute simplicity of manner and winningness of smile when his face lights up, that are charming . . .', William James about Ernst Mach, letter to his wife of 2 Nov. 1882 quoted from Joachim Thiele [1978], p. 169. Of course, it is difficult for a dweller in the slums of late twentieth century philosophy of science to guess what an intelligent man may or may not assume.

[5] Or else it shows that Popperians, being engulfed by the banality of their own 'philosophy' have completely lost their sense of perspective.

Now despite this disadvantage the theory of relativity might still be welcome as a step in the right direction were it not for the fact that it has been 'growing more and more dogmatical' (first reason). Was this the case? Was there really an increase of dogmatism, or what Mach regarded as dogmatism, in physics, in connection with the theory of relativity? Not on Einstein's part but certainly on the part of Planck, Mach's arch enemy.[6] Planck, after some wavering, defended a reality of the kind Mach and Boltzmann had rejected, he regarded relativity as a step towards this reality, made himself many contributions towards the development and extension of the theory[7] and so furthered its dogmatic interpretation. I conjecture that this explains Mach's criticism and the harshness of it but it also shows that ***the criticism does not touch Einstein's achievements.*** It concerns only their misuse for dogmatic philosophical purposes.[8]

---

[1] Mach [1922], p. 255.

[2] *Op cit.*, p. 26.

[3] Note that Zahar cannot explain the role of experiments in Mach's objections: one does not need experiments to show that a theory is metaphysical.

[4] Mach [1896], p. 228.

[5] There are various hypotheses concerning the 'experiments'. Thus Herneck conjectures that Mach objected to the assumption of the constancy of c and that the 'experiments' dealt mainly with these matters: D. K. Heller [1964], p. 142. According to Herneck, Philipp Frank gave Dingler part of the responsibility for Mach's attitude. (*Physikalische Blätter*, **xvii**, (1961), p. 276). However the fact that Mach emphasised the need to combine psychological, physiological, physical as well as biological investigations both in his early and in his late writings and that he deplored a purely physical approach to the problem of space seems to favour my interpretation: relativity was OK especially as it introduced the interdependence of spatial and temporal aspects but it did not go far enough and it became a hindrance to research if interpreted dogmatically, in the manner of Planck.

[6] *Cf.* the rather acrimonious exchange between Mach and Planck, republished in Toulmin [1968]. Although the tone of learned exchange was at the time a little less vapid than it is today, this was the only polemic Mach ever published.

[7] *Cf.* the account and the literature in Stanley Goldberg [1972].

[8] Similar remarks apply to Mach's rejection of atoms. First a little myth should be laid to rest. Stefan Meyer of the Institute for Radium Research at the University of Vienna reports and many writers have repeated that Mach, having been shown a spinthariscope

9 Is Einstein's work influenced by Mach? I think it is. However the influence becomes clear only if one disregards technical details and looks at the wider picture. We then notice that Einstein strives for the most direct expression of phenomena: electric induction depends only on the relative motions of the systems concerned while the aether theories treat the effects of the motion of a magnet (relatively to the aether) differently from the effects of the motion of a closed electric circuit.[1] Lorentz explains null results approximately and using a variety of rather doubtful hypotheses[2] while Einstein regards them as basic facts.[3] In his investigations of the structure of radiation Einstein does not start from basic theory, as Planck did, but from uncomprehended but empirically adequate laws and develops their 'ontological' consequences. 'The key to his reasoning' writes Martin Klein on this procedure[4] 'was his reversal of Planck's procedure. Instead of trying to derive the distribution law from some more fundamental starting point, he turned the argument around. Planck's law had the solid backing of experiment; why not assume its correctness and see what conclusion it implied as to the structure of radiation?' He plays 'theoretical pictures'[5] (his term—most likely taken from Boltzmann whom he read carefully[6]) off against each other to find their limits.[7] Like Mach, Einstein is mainly interested in 'theories of principle',[8] *i.e.* in theories which cover a wide variety of phenomena without assumptions about details of their constitution (this also

---

said 'jetzt glaube ich an die Atome', *Sitzungsberichte der Oesterreichischen Akademie der Wissenschaften—Mathematisch-Naturwissenschaftliche Klasse*, Abt 2a, Band 159, Heft 1/2, 5. Heinz Post has pointed out that Mach was a kind person, that Meyer was very enthusiastic and rather deaf and that he most likely misunderstood the kind remarks Mach made in such a personally difficult situation. And who would think that Mach who had not been impressed by some very powerful arguments (Brownian motion, for example) should fall for such a flimsy exhibition? Secondly, Mach's objections were firmly grounded in some very real difficulties of the atomic theory which were often forgotten in view of apparent successes. Finally, the nineteenth century atom has indeed disappeared and has been replaced by a much more complex entity, constituted out of relations between well-defined measurements and no longer in conflict with Mach's ideas.

[1] 'On the electrodynamics of Moving Bodies' quoted from Einstein and others [1923] p. 37.

[2] *Cf.* the articles by Lorentz in Einstein and others [1923].
'I have not been able to obtain for the equations referred to moving axes *exactly* the same form as for those which apply to a stationary system.' Lorentz [1952], p. 230.

[3] 'Einstein simply postulates what we have deduced, with some difficulty, and not altogether satisfactorily, from the fundamental equations of the electromagnetic field. By doing so, he may certainly take credit for making us see in the negative result of experiments like those of Michelson, Raleigh and Bruce, not a fortuitous compensation of opposing effects, but the manifestation of a general and fundamental principle.' Actually, Einstein did not 'make us see it' that way, for there was no explanation, he simply took this situation as his starting point: 'Wir wollen diese Vermutung zur Voraussetzung erheben'—Einstein and others [1923], p. 26. English translation *op. cit.*, p. 38.

[4] Klein [1964], p. 9. *Cf.* also T. S. Kuhn [1978], 180 f.

[5] *Cf.* the first page of his paper on the photoelectric effect reprinted *e.g.* in ter Haar [1967], p. 91.

[6] *Cf.* the references in Seelig's biography.

[7] Mach was well aware of the fact that different theories could cover some domains equally well. *Cf.* also his remarks on analogues in science his [1917], p. 220 ff.

[8] *Cf.* the autobiographical notes in Schilpp (*ed.*) [1949], esp. p. 53.

establishes a link with Duhem). Einstein's remarks on geometry are in perfect agreement with what is found in the relevant chapters of ***Erkenntnis und Irrtum***[1] and in earlier writings.[2] As regards method, Einstein emphasises the great difference between the work of the scientist and the work of the epistemologist (philosopher of science): 'The external conditions' he writes[3] 'which are set [for the scientist] by the facts of experience do not permit him to let himself to be too much restricted in the construction of his conceptual world, by the adherence to an epistemological system. He, therefore, must appear to the systematic epistemologist as a type of unscrupulous opportunist.' 'Yes, I may have started it' he tells Infeld about a new fashion in physics—'but I regarded these ideas as temporary. I never thought that others would take them so much more seriously than I did.'[4] Or, in a lighter mood: 'A good joke should not be repeated too often.'[5] Compare this with Mach's 'the schematisms of ***formal*** logic and also of ***inductive*** logic are not of much use, for the intellectual situations are never repeated in exactly the same way. But the examples of great scientists are very instructive . . .',[6] the already quoted passage 'there is no need to turn ***this passing point of view*** (*i.e.* Mach's own suggestions) into a system for life whose slaves we then become'[7] as well as with Mach's modest and somewhat mocking attitude towards his own 'philosophy'[8] and you will see that Einstein and, as a matter of fact, all creative scientists are much closer to Mach than to, say, Popper. Mach, Einstein, Bohr use philosophy as an instrument of research that guides research but is also transformed by it while philosophers pontificate on a process they do not understand and would not know how to change. Small wonder they want to persuade us that Mach's ideas are as vapid as their own.

University of California, Berkeley
Eidgenössische Technische Hochschule, Zürich

REFERENCES

BOLTZMANN, L. [1905]: *Populaere Schriften*, Leipzig.

CAJORI, F. (*ed.*) [1962]: *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World*, Berkeley: University of California Press.

CARNAP, R. [1928]: *Der Logische Aufban der Welt*; all page references are to the reprint, Hamburg 1974.

CLARK, R. W. [1971]: *Einstein.* New York.

---

[1] *Cf.* ch. 1 of Einstein's [1922] with pp. 353 ff. of Mach [1917].

[2] *Cf.* Mach [1866], his set essay on the development of our ideas of space, esp. p. 232. *Cf.* also Mach's note concerning Minkowski's 'Raum und Zeit' in the new (1909) edition of 'Die Geschichte und die Wurzel des Satzes von der Erhaltund der Arbeit' with reference to the relevant chapters of [1917]: 'Space and time are not regarded as self contained entities but as forms of dependence of the phenomena. I therefore move towards the principle of relativity which is also asserted (festgehalten) in 'Mechanik' and 'Waermelehre'—p. 60.

[3] Schilpp (*ed.*) [1949], p. 683f.

[4] Quoted from R. W. Clark [1971], p. 340.

[5] Philipp Frank [1948], p. 261.

[6] Mach [1917], p. 200. The same applies of course to the various theories of rationality which are of no use to the scientist but only provide philosophers with material for endless quarrels. 'Die Wissenschaft ist kein Advokatenkunststueck' says Mach [1917] and he is certainly right.

[7] Mach [1896], p. 226.

[8] Mach [1922], p. 39.

CURD, M. S. [1978]: *Ludwig Boltzmann's Philosophy of Science.*

DINGLER, H. [1924]: *Die Grundgedanken der Machschen Philosophie*, Leipzig.

EINSTEIN, A. [1922]: *The Meaning of Relativity*, Princeton.

EINSTEIN, A. [1936]: 'Physics and Reality', *Journal Frankl. Instit.* **221**; all page references to the reprint in *Ideas and Opinions*, New York, 1954.

FEYERABEND, P. K. [1978]: *Science in a Free Society.* London.

FRANK, P. [1948]: *Einstein, His Life and Times.*

GOLDBERG, S. [1972]: 'Max Planck's Philosophy of Nature and his Elaboration of the Special Theory of Relativity', *Historical Studies in the Physical Sciences*, 3 Philadelphia.

KUHN, T. S. [1978]: *Black-Body Theory and the Quantum Discontinuity.*

LORENTZ, H. A. [1952]: *The Theory of Electrons.* New York.

MACH, E. [1866]: *Zeitschrift für Philosophie und philosophische kritik.*

MACH, E. [1896]: *Populaerwissenschaftliche Vorlesungen.* Leipzig.

MACH, E. [1917]: *Erkenntnis und Irrtum.*

MACH, E. [1922]: *Analyse der Empfindungen.* Jena.

MACH, E. [1933]: *Die Mechanik*, 9th Edition. Leipzig.

SCHILPP, P. A. (*ed.*) [1949]: *Albert Einstein: Philosopher Scientist.* New York.

TOULMIN, S. [1968]: *Physical Relativity.* Harper Torchbooks.

ON THINGS AND CAUSES IN SPACETIME*

Introduction
1 *Things in Spacetime*
2 *Causes in Spacetime*

INTRODUCTION

Graham Nerlich's admirable article 'What can geometry explain?' [1979] persuades me that we should take the General Theory of Relativity to have reduced 'gravitation to spacetime curvature' (p. 81). I agree with Nerlich that spacetime is an entity, irreducible to the spatiotemporal relations between the things and events in it; and that the variable curvature of its geodesics, induced by the distribution of matter, explains geometrically what we see as acceleration caused by gravitational forces.

From these truths, however, Nerlich has drawn two further, false, conclusions. One is that we must give up 'the space and time language of enduring continuants' (p. 81); the other is that a serious puzzle is generated by 'the possibility of causal intervention by us to change the gravitational field' (p. 82). I take these points in turn, since the first is the more important and less obvious of the two and, when it is dealt with, the second can be disposed of quite easily.

1 THINGS IN SPACETIME

I take it that by 'continuants' Nerlich means what have been called 'substances' as opposed to events. I shall call them 'things', and I take them to differ from temporally extended events in lacking temporal parts. The question then is why accepting General Relativity as Nerlich construes it should force us to deny the existence of things in this sense. To tackle the question I must first elaborate a little on the distinction between things and events.

---

* This note was written during my tenure of a Radcliffe Fellowship, for which I am indebted to the Radcliffe Trustees.

Examples of temporally extended events are: wars, meals, avalanches, eclipses, explosions, nuclear reactions and the expansion of the Universe. All these events take time, and none is wholly present at any one time. An instant of time indeed contains no part of such events; it merely separates temporal parts of them, as an internal surface separates the spatial parts of a thing. It takes a stretch of time to contain any temporal part of an event. An event can be said to be present at a time only in the sense in which I am present at every point of space within my body.

Things, on the other hand, are wholly present at every time at which they exist at all. Examples of things are: nations, people, omelettes, mountains, planets, bombs, atoms and galaxies. None of these things has temporal parts in the way it has spatial parts. It is, for example, nonsense to say that only a youngish part of Sir Edmund Hilary climbed a rather older part of Everest. Hilary himself climbed Everest itself, and both parties were wholly present during all temporal parts of that historic event.

The distinction between things and events is best expressed by different parsings of sentences attributing change to them. Suppose an entity a has some property F at time t which it loses by time t^*, acquiring instead a contrary property G. F and G might be different colours, or temperatures, or electric potentials: any respectable and changeable property, physical, biological or mental. I exclude factitious properties like Goodman's [1955] 'grue', though I shall not try here to defend any particular criterion for distinguishing them from real properties. Then the facts are that

$$a \text{ is } F \text{ at } t \tag{1}$$

and

$$a \text{ is } G \text{ at } t^*. \tag{2}$$

If a is an event, the time reference is taken to be part of the subject, thus:

$$a\text{-at-}t \text{ is } F \tag{3}$$

and

$$a\text{-at-}t^* \text{ is } G. \tag{4}$$

That is, we suppose that different temporal parts of a, containing respectively the times t and t^*, are F and G. This is like different spatial parts of a thing being at different temperatures. The whole thing is not at any of the temperatures; likewise, the whole of a is neither F nor G unless, contrary to our hypothesis, all parts of it are. So a itself does not in fact *change* in respect of F-ness; there are merely differences in this respect between different entities, namely, temporal parts of a containing respectively t and t^*. And these in turn do not change: each is either wholly F or wholly G, or itself divides into temporal parts which are wholly F or wholly G. Events therefore do not change at all, although they may *be* changes, namely changes in things.

If a is a thing, it has no temporal parts and the time reference is not taken to be part of the subject in (1) and (2). a itself is the subject, and F is construed as a relation between *a* and the times at which it is *F*; likewise *G*, so that (1) and (2) become

$$F(a, t) \tag{5}$$

and

$$G(a, t^*). \tag{6}$$

That *a* changes between *t* and t^* appears in the fact that *a* has incompatible relations *F* and *G* to these different times, *i.e.* relations such that nothing could have them both at the same time. Now it has been said (not by Nerlich) that this is an inadequate account of change, because (5) and (6) are true at all times. But so of course are (1) and (2). The assumption behind the complaint is that as well as what McTaggart [1908] called the '*B*-series', in effect of times in the sense of dates (1980, 29 May, 2 p.m., *etc.*), there is an *A*-series, of presentness and of the greater and less degrees of pastness and futurity; and that change consists in the changing relations of the elements of the two series, in this case from

$$a \text{ is (now) } F \tag{7}$$

when *t* is the present time and t^* is still future, to

$$a \text{ is (now) } G \tag{8}$$

when t^* becomes present and *t* has become past. But seventy years on, McTaggart's proof really should have sunk in, that this kind of change (call it 'McTaggart change') involves inescapable contradiction. There is no *A*-series; there are in reality no such nonrelational properties as being present, or being more or less past or future. There are only times in the sense of dates; and these are times, as opposed say to places, because they are ordered by an 'earlier than' relation of which, *pace* McTaggart, temporal sense can be made without appeal to the existence of a corresponding *A*-series. (This needs showing, but here is not the place to show it, since it is not a point at issue between Nerlich and myself.) That being so, all there can be to change is what is captured in (5) and (6).

Once the itch for McTaggart change has been cauterised, (5) and (6) really ought not to be contentious. To say that *F* is a relation is only to say that both the contexts

'... is *F* at *t*'

and

'*a* is *F* at ...'

are transparent (*i.e.* it doesn't matter to the truth of (1) how either *a* or *t* is referred to), and that for (1) to be true, *a* and *t* must both exist. (This last condition is of course compatible with taking times to be constructions out of temporal relations between things and events; one need not, like Newton, Nerlich and myself, believe in absolute spacetime.) Now both these claims are plainly true of all the common and important changeable properties of things: being red, massive, at 100°C, negatively charged, metallic, *etc.* The only difference relativity makes is that *t* becomes, not a unique cross-section of Newtonian spacetime, but a position on the timelike world-line of each thing.

But Nerlich says that General Relativity as he interprets it could not apply to things, so that to accept it is to deny that there are any. Were that so, General Relativity would have us regard the concept of a thing as a hangover from classical physics; a fiction which, however convenient and engrained in our thought and discourse, must be dispensable from a true description of the world. My own view is that the concept is both important and indispensable: important amongst other things because without it we could not apply concepts of moral or legal responsibility; indispensable because we cannot account for our perception of time itself unless we are things rather than events. I shall not try to argue these propositions here, although if true they show that it is General Relativity, not things, which would have to go should the two prove inconsistent. But in fact they are perfectly consistent, so far as I can see. The question is, why does Nerlich think otherwise?

Nerlich does not make his argument explicit, but I take it to be roughly as follows. Explaining geometrically what appears in space and time as things accelerating under gravitational forces reduces them to four-dimensional spacetime trajectories. Trajectories are events, because they have temporal parts that are essential to the explanation. It is from the spacetime curvature, which varies from part to part of a trajectory, that apparent movements of the thing in space are derived. The reality then is spatiotemporal events, which only appear as things under the pre-relativistic concepts, mistakenly separated, of space and time.

If anything like this is the argument (and I perceive no other), it seems to me fallacious. The fallacy lies in failing to distinguish merely possible trajectories, which indeed exist as parts of spacetime, from the things which may or may not have them as actual trajectories. I readily admit that spacetime is a whole of which possible trajectories are parts, and that these in turn have temporal parts, thus counting, if anything, as events rather than things. But when we talk of a thing's actual trajectory the matter is different. Its trajectory is simply a property the thing has. Because we can distinguish temporal parts in the region of spacetime the thing occupies, it does not follow that the thing itself has temporal parts.

Consider the familiar pre-relativistic fact that things, including people, share with events the property of duration. Churchill had a span of years, just as the Second World War did. But it does not follow that to each decade within that span there corresponds a different temporal part of Churchill. (If it did follow, it would be surprising that the concept of a thing has lasted so long; certainly it would not take General Relativity to dispose of it.) There are indeed various temporal parts of Churchill's *life*, by which I mean a generalisation of his trajectory to take in all his changeable properties and relations, not just his spatial location. His account of one of these parts is quite properly called 'My early life'. But not 'Early me'. It is obvious from their respective identity conditions that a man is not the same entity either as his life or as any temporal part of it. A thing *a* is not *a*-at-*t*, nor *a*-from-*t*-to-*t**, nor any whole comprised of such parts. We can indeed define such wholes, and such parts, given a sufficient knowledge of *a* and of the criteria which determine its persisting identity through time. But these are logical constructions out of *a* and times, not vice versa. It takes real physics and metaphysics, not just typography, to establish that a temporally extended entity is an event, that it has independently identifiable temporal parts of which it is a whole. The present question is whether General Relativity, on Nerlich's interpretation, does indeed establish this of what we have for so long taken to be things.

I cannot see that it does. For a start, General Relativity explains very little. Most of a thing's changeable properties have nothing to do with gravity, and a theory of gravity gives no reason at all to parse change in them according to (3) and (4) rather than (5) and (6). General Relativity does not reduce things to events; at most it reduces their gravitational behaviour to properties of their trajectories. But even if we took things to have only inertial and gravitational properties, they could still not be regarded as constructions out of trajectories. On the contrary, a thing's trajectory, like Churchill's life, can itself only be identified by appeal to the persisting identity of the thing whose trajectory it is. Not every timelike path in spacetime, whose parts are all occupied by events, is the trajectory of a thing. And only those that are, which we need thing-concepts to pick out, represent the movement or acceleration which it is the business of a gravitational theory to explain. We have seen that only things change. In particular, only things travel and accelerate and so call at all for gravitational explanation, be it causal or geometrical.

The geometrical explanation of a thing's gravitational acceleration indeed invokes the curvature of temporal parts of its trajectory. So in accepting this explanation we commit ourselves to the trajectory and its parts as entities independent of the thing that has them. Likewise in explaining a photon's progress through refracting media by its taking the fastest path between its entry and exit points, we commit ourselves to that path, which is also a trajectory and has temporal parts. But that does not make it sensible to credit the photon with temporal parts, or to suppose that this explanation reduces the photon to the path it follows. No more is it sensible to take the geometrical explanation of gravity as reducing things to the event-like regions of spacetime whose status as trajectories itself depends on all their parts being occupied by one and the same thing. On the contrary, to dispense with the concept of a thing is to dispense also with the concepts of travel and acceleration, without which gravity itself has nothing to explain. So far therefore from the General Theory of Relativity forcing us to banish things from our ontology, there would be no gravitational phenomena left for it to explain if we did.

2 CAUSES IN SPACETIME

A thing's spacetime trajectory is one of the thing's properties, but not one of those schematised in (1) and (2). It already comprises all the spacetime points the thing is ever at, so there is no sense to supposing it different at different times. Unlike mass, colour, charge, *etc.*, trajectories are not changeable properties. So they are not, as changeable properties are, really relations between things and various times. A trajectory is just what it seems: a single, nonrelational property of a thing.

A thing's trajectory then cannot change, but it might of course have been other than it is. In particular, we might have caused it to be so; for example, by pushing the thing out of the path it would otherwise have followed. Now since the distribution of matter determines a gravitational field, which in turn determines the local structure of spacetime, it follows that by pushing things around I can affect this structure, albeit only very slightly and only within the forward light cone, since effects on spacetime, like all effects, must be later than their causes. Out of this seemingly innocuous fact Nerlich generates the following puzzle:

> If spacetime structure were affected by an action then it would be changed by it. This ought to mean that, at one time, the curvature of a spacetime region has one value and, at another time, another value. But that is strictly unintelligible unless we erect a further dimension of time within which spacetime could undergo such a change. (p. 82).

This is in fact a quite spurious puzzle, generated by confounding being an effect with being a change. The world does of course contain changes, namely changes in things; and like other events, most if not all changes are effects. Consider for example the change in *a* from being *F* at *t* to being *G* at t^*. Suppose there to be a cause *e* of this change, without which *a* would have remained *F* from *t* to t^*. Then the facts of this change are:

$$F(a, t) \tag{5}$$

and

$$G(a, t^*). \tag{6}$$

The counterfactual which expresses the fact that *e* is the cause of this change is:

$$\text{Had } e \text{ not occurred, } a \text{ would have been } F \text{ at } t^*. \tag{9}$$

If one confounds the factual statement of the change with the counterfactual statement of its cause, one is apt to think *e* causes *a* to change from being *F* at t^* to being *G* at the very same time; and this indeed is 'strictly unintelligible unless we erect a further dimension of time within which [*a*] could undergo such a change'. But we can forgo the effort of this erection: this supposed change simply does not exist, and there is no reason to try and make it intelligible. Either *e* occurs and $G(a, t^*)$, or it doesn't and $F(a, t^*)$. The world neither way contains both $G(a, t^*)$ and $F(a, t^*)$, nor therefore any change along some further time dimension from one to the other.

The spacetime which concerns Nerlich is not of course a thing. It has temporal parts and so, like the events it contains, neither it nor its parts are susceptible of change at all. But like other events, they can, as Nerlich observes, be affected by human action. These effects are expressed, like the effects on *a* above, by counterfactuals: had the action not occurred, the structure of spacetime would have been different. So it would: but either the action does occur and there is one structure, or it doesn't and there is another. The world neither way contains both structures, nor therefore a change along Nerlich's further time dimension from one to the other.

There is no change in spacetime at all, because spacetime and its parts are not things. But there is causation in it nonetheless: of spacetime structure by the distribution of matter in it, and of the trajectories of things by spacetime structure.

D. H. MELLOR
University of Cambridge

REFERENCES

Goodman, N. [1955]: *Fact, Fiction, and Forecast.* Cambridge, Mass.: Harvard University Press.

McTaggart, J. E. [1908]: 'The unreality of time', *Mind*, **18**, pp. 457–74.

Nerlich, G. [1979]: 'What can Geometry Explain?' *British Journal for the Philosophy of Science*, **30**, pp. 69–83.

Brit. J. Phil. Sci. 31 (1980), 289–298 *Printed in Great Britain*

Review Article

PHENOMENOLOGY AND THE INFINITE IN MATHEMATICS*

1 *Introduction*
2 *Kaufmann's Phenomenology*
3 *Kaufmann's Treatment of Arithmetic*
4 *Kaufmann's Denial of the Actual Infinite*
5 *General Conclusions*

1 INTRODUCTION

Felix Kaufmann (1895–1949), though a follower of Husserl, nonetheless regularly attended the meetings of the Vienna Circle during its heyday. His original training was in the philosophy of law, and he was for a while an unsalaried *Privatdozent* at the University of Vienna. He did not, however, follow a full-time academic career, but worked in business, becoming manager of the Austrian branch of the Anglo-Iranian Oil Company. This did not lead him to abandon his philosophical interests, and he had published no less than three books on the philosophy of law before making his debut in the philosophy of mathematics in 1930.

The present volume contains English translations of Kaufmann's book on philosophy of mathematics (*The Infinite in Mathematics and its Elimination*, first published 1930), and of two articles which cover much the same ground. There is also a useful introduction by Ernest Nagel.

The first of the two articles: 'Remarks on the Controversy about the Foundations of Logic and Mathematics' appeared in *Erkenntnis* in 1931. The second 'Questions of Logical Principle in the Investigation of the Foundations of Mathematics' was intended for Husserl's *Jahrbuch für Philosophie und phänomenologische Forschung*, but was never published and appears in print for the first time in the present volume.

Kaufmann's book therefore appeared at the end of a decade of exciting developments in the foundations of mathematics, but just before Gödel's classic 1931 paper: 'On Formally Undecidable Propositions of *Principia Mathematica* and related systems.' Kaufmann's lack of knowledge of Gödel's incompleteness theorems leads him, as we shall see, into a number of errors, but it is very far from destroying completely the interest of his work. Kaufmann's position can only be properly understood against the background of the then current controversies in the philosophy of mathematics, and we will next try to sketch very briefly the nature of these.

---

* Review of KAUFMANN, F. [1978]: *The Infinite in Mathematics*. Edited by Brian McGuiness with an Introduction by Ernest Nagel. Translated from the German by Paul Foulkes. Vienna Circle Collection. Dordrecht: D. Reidel. D.fl. 85. Pp. 235.

[1] I am very grateful to Dr M. Machover and Dr M. L. G. Redhead for reading an earlier draft of the review, and suggesting improvements.

The great interest in philosophy of mathematics in the 1920s was undoubtedly provoked by Brouwer and Weyl. Brouwer first elaborated his intuitionistic philosophy of mathematics in the period 1907–9. However, this early work did not have a big impact. This was partly because his early papers were written in Dutch, and not translated into a more generally accessible language. Indeed Brouwer's first criticism of the law of the excluded middle (his 1908 paper on 'The Unreliability of the Logical Principles') was translated from Dutch for the first time for the 1975 edition of Brouwer's Collected Works, where it appears in English. Another point is that Brouwer's early papers are mainly a philosophical critique of standard mathematics, and do not contain any mathematical developments. In 1918, however, Brouwer began publishing a series of papers in German in which he developed intuitionistic mathematics. The title of the first of these (A Foundation for Set Theory independent of the law of the excluded middle) clearly indicated Brouwer's general programme. He wanted to build up a new, constructive, mathematics which did not use those principles of classical logic and mathematics which he regarded as invalid—particularly of course the law of the excluded middle.

In 1918 Weyl, working independently of Brouwer, published a booklet 'The Continuum' in which he attempted to give a constructive development of analysis. His original version of constructivism differed from Brouwer's, but in 1921 he declared: 'I now give up my own attempt and join Brouwer.'

This flowering of constructivism was very displeasing to Hilbert. Hilbert was throughout his life an opponent of constructivism, which he regarded as harmful and restrictive. In his youth he had opposed the constructivist views of Kronecker, and in 1921 he seems to have been particularly annoyed that one of his favourite pupils (Weyl) should have espoused constructivism. He therefore abandoned the work in applied mathematics and physics which had occupied him since 1902, and set out to defend mainstream mathematics.

Hilbert's programme was to formalise the standard systems of mathematics (arithmetic, analysis, and set theory), and then to prove these systems consistent using metamathematical reasonings involving only finitary (and hence indubitable) principles. In this way he hoped to justify classical mathematics, and put it on a firm basis. The programme in its original form was shown to be impossible to carry out by Gödel in 1931.

To sum up then. The philosophy of mathematics was dominated in the 1920s by the opposition between Brouwer's proposals for a new intuitionistic mathematics, and Hilbert's defence of classical mathematics. In his book, Kaufmann attempts to steer a middle course. He accepts the law of the excluded middle, but nonetheless agrees with Brouwer in rejecting some parts of Cantorian set theory as inadmissible. Kaufmann argues for this middle course on the basis of Husserl's philosophy. In the next section we will briefly examine some points in this philosophy, and how Kaufmann applies them to the foundations of mathematics.

2 KAUFMANN'S PHENOMENOLOGY

Kaufmann relies principally on two doctrines borrowed from Husserl's phenomenology. The first of these is a kind of Aristotelian view of universals.

Properties exist only as instantiated in material objects. The mind can, by a process of abstraction, consider properties separate from objects. But properties do not really have an existence independent of physical objects. The Platonic supposition that they do is inadmissible metaphysics. As Kaufmann puts it (pp. 20–1)[1]:

> . . . they (*i.e.* the properties of a corporeal thing) are not isolated, but only capable of being considered in isolation. The possibility of thus isolating them in thought is, however, based on free variability within certain limits, . . . For example, colour and extension belong inseparably together, but nevertheless can be considered each on its own, because any arbitrary colour is compatible with any arbitrary extension. If, however, we hypostasize this isolability into independent existence, then we become involved in boundless metaphysical speculation.

This point of view leads Kaufmann to allow only 'properties of material objects', and to reject all talk of 'properties of properties', 'properties of properties of properties', *etc.*, except in so far as such discourse can be translated into an equivalent form in which mention is only made of properties of corporeal things, and not of properties of other properties. As Kaufmann himself says (p. 23):

> . . . wherever people talk of 'properties' of properties, 'properties' of 'properties' of properties and so on, the word 'property' has no uniform meaning: such propositions therefore have a sense only insofar as they are 'translatable' into other propositions in which the word 'property' is not used in this twofold manner.

This point of view leads at once to the rejection of the extended functional calculus as Kaufmann goes on to observe. It also leads to a truncation of Cantorian set theory. Cantor introduces his exponential hierarchy of alephs in the following manner. He begins with the set of natural numbers (N say) which has cardinality $\aleph_0$. He then considers the set of all subsets of $N =$ the power set of $N = PN$. This has cardinality $2^{\aleph_0}$, and Cantor uses the diagonal argument to show that $2^{\aleph_0} > \aleph_0$. Similarly PPN has cardinality $2^{2^{\aleph_0}} > 2^{\aleph_0}$ and so on. Just as Kaufmann rejects properties of properties, properties of properties of properties *etc.* except as a *façon de parler*, so analogously he must reject sets of sets, sets of sets of sets, *etc.* except in the same limited sense. Thus he cannot allow PN, PPN, . . . and $2^{\aleph_0}$, $2^{2^{\aleph_0}}$, ... Generally speaking Kaufmann's view of universals leads to a considerable truncation of classical mathematics—indeed to a greater truncation than he himself realises. This is a point to which we will return.

The second main doctrine which Kaufmann takes from phenomenology is one which he could equally well have obtained from the Vienna Circle. It is that the task of philosophy is not 'speculative construction', but 'clarification of thought through reflection'. Correspondingly (p. 3):

> The task of a philosophy of mathematics . . . consists in bringing out with full clarity, the sense of mathematical propositions, concepts and method.

---

[1] All quotations from Kaufmann are taken from the book under review.

It is interesting to note then that phenomenologists and logical positivists, whatever their differences, are united in condemning metaphysics.

3 KAUFMANN'S TREATMENT OF ARITHMETIC

Kaufmann begins by rejecting the Frege–Russell definition of number. He believes that the notion of number is simpler and more basic than that of set. As he says (p. 69):

> . . . in defining the concept of number we do not require the concept of a set.

We obtain the concept of number by abstracting from the concrete process of counting, just as, to use his earlier example, we obtain the concepts of colour and extension by abstraction from our concrete experience of material objects. Despite his emphasis on counting, Kaufmann denies the Kantian view that there is a close connection between number and time. Thus he says (p. 71):

> . . . in trying to find an exact definition of natural numbers . . . we must use neither the concept of time nor that of a totality of things or of a property of a complex (set, class and so on) . . . . Numbers turn out to be logical abstractions of the counting process conceived as being capable of being continued without bounds.

This view makes the notion of ordinal number more basic than cardinal number, and Kaufmann disputes the Russellian thesis that the opposite is the case. However, Kaufmann does stress that it is fundamental to the number concept that counting a totality in different orders always yields the same result. He goes on to argue that his definition of number as a logical abstraction of the counting process is equivalent to Peano's axiomatic approach (p. 77):

> . . . our definition of natural numbers is not only equivalent to Peano's axiomatic system but agrees with it as regards procedure, . . .

Kaufmann then tries to use his informal definition of number to show that the Peano axioms are consistent and complete, and that every problem of arithmetic is decidable. He sums up what he has claimed to have shown as follows (pp. 179–80):

> The consistency or completeness of this system of determinations defining the series of natural numbers is simply the consistency or completeness of arithmetic (including analysis), since the latter is completely defined by the concept of the natural number series. In particular this implies that there can be no problem of arithmetic that is in principle undecidable. For undecidability is merely indeterminacy within the presuppositions, as is easily seen if we free ourselves from the non-concept of the possibility of infinitely many steps of demonstration.

Of course Gödel's incompleteness theorems cast doubt on these claims. Gödel's first theorem shows that the Peano axioms[1] are incomplete, and that there exist arithmetical propositions which are not decidable on the basis of these axioms. The application of Gödel's second incompleteness theorem is not so immediately clear, since Kaufmann's proof of consistency is informal in character. However, Kaufmann's informal procedures are certainly weaker than Peano arithmetic. So his informal proof, were it correct, could be formalised within Peano arithmetic, and would therefore, by Gödel's second incompleteness theorem, show Peano arithmetic to be inconsistent. Thus, if we assume Peano arithmetic is consistent, Kaufmann's informal consistency proof cannot be valid.

---

[1] When speaking of the Peano axioms, or Peano arithmetic, I shall throughout mean 'the first-order Peano axioms' or 'first-order Peano arithmetic'. Kaufmann must also understand Peano arithmetic in this sense, since, as we have seen, he rejects the extended functional calculus.

Moreover, Kaufmann's claim to have obtained a system as strong as Peano arithmetic merely by abstracting from the counting process seems to me doubtful. Frege has a well-known argument (*cf.* Foundations of Arithmetic, §5 pp. 5–7 and §7 pp. 9–11) to the effect that equations involving large numbers *e.g.* $135664+37863 = 173527$ are not based on simple intuitive or empirical counting procedures. We could not in practice verify such an equation directly by counting, but are convinced of its truth through our grasp of a theoretical schema (decimal notation). If this line of thought is correct, abstracting from concrete counting would give us only small numbers and their arithmetic. To develop the arithmetic of large numbers, and, still more, to introduce Peano arithmetic we have to go beyond concrete counting procedures.

Let us now turn to Kaufmann's critique of Brouwer. Of course Kaufmann is in partial agreement with Brouwer. As he says (p. 58):

> Brouwer's critique is valid and important, insofar as it is directed against the doctrine of actual infinity in general and the non-denumerable infinite in the sense of Cantor's set theory in particular; . . .

We shall examine this matter in the next section.

Kaufmann, however, rejects Brouwer's account of how numbers are created from the basal intuition. To begin with he feels that Brouwer is wrong to follow Kant in basing mathematics on intuitions of time (p. 58):

> . . . we should reject that part of Brouwer's doctrine which is essentially based on the introduction of the concept of time into mathematics.

At a more fundamental level, Kaufmann feels that Brouwer's view of numbers and other mathematical entities as created out of pure subjective cognition is untenable. Here Kaufmann is again following the principles of phenomenology. According to these, any act of cognition must be cognition of some object which exists independently of being cognized. Abstract entities really exist only as instantiated in material objects. Thus to imagine such entities created out of pure cognition (as Brouwer does) must be wrong. This is how Kaufmann himself puts it (pp. 56–7):

> . . . Brouwer's view . . . that there is something cognizable that is created only by being cognized, . . . runs counter to the nature of cognition. For, as we saw . . ., all cognition presupposes an object that must be thought of as existing independently of its being cognized.
>
> . . . mathematical propositions are discovered not invented. . . .

Brouwer argued that the law of the excluded middle implies that all mathematical questions are decidable. Since he believed there was not a shred of evidence that this was really the case, he argued that the law of the excluded middle should be rejected. Now, as we have seen, Kaufmann argued erroneously that arithmetic and analysis were complete and decidable, and, reversing Brouwer's argument, he went on from this false premise to argue in favour of the law of the excluded middle. As he puts it (p. 180):

> Moreover, it is the completeness of the system that allows the 'unrestricted application of the law of excluded middle' in mathematics.

Obviously this argument can no longer be accepted in the light of Gödel's incompleteness theorems. But what of his more philosophical critique of Brouwer? Here, as elsewhere, Kaufmann is appealing to the phenomenological principle that properties and abstract entities do not have an existence independent of corporeal things, but really exist only instantiated in material objects. This is a very plausible general philosophical view, but, unfortunately, as I shall next argue, it leads to more of a truncation of ordinary mathematics than Kaufmann realises.

Let $P(a)$ ascribe a property P to a material object a. Now the phenomenological principle states that properties like P do not have an existence independent of material objects like a. This seems to imply that we should not consider a Platonic domain of properties P, and correspondingly should not quantify over properties, or indeed over other abstract objects such as sets, numbers, *etc.* Thus the phenomenological principle seems to imply that, in the quantifiers $(\forall x)$ $(\exists x)$, x should range only over material objects (or possibly psychological objects such as thoughts), but not over abstract entities. Let us now examine how much of ordinary arithmetic and theory of numbers can be retained if we accept this prescription.

Certainly some numerical statements can still be made. Perhaps the easiest way to do so is to introduce the numerical quantifiers $(\underset{0}{\exists} x)$, $(\underset{1}{\exists} x)$, $(\underset{2}{\exists} x)$, . . ., $(\underset{n}{\exists} x)$, . . . where $(\underset{n}{\exists} x)\, P(x)$ is supposed to mean: 'there are exactly n x such that $P(x)$'. These quantifiers can be defined as follows:

$$(\underset{0}{\exists} x)\, P(x) =_{\text{def}} \neg\, (\exists x)\, P(x)$$

$$(\underset{1}{\exists} x)\, P(x) =_{\text{def}} (\exists x)\, (P(x) \wedge (\forall y)\, (P(y) \rightarrow y = x))$$

$$(\underset{2}{\exists} x)\, P(x) =_{\text{def}} (\exists x)\, (\exists y)\, (P(x) \wedge P(y) \wedge x \neq y$$
$$\wedge\ (\forall z)\, (P(z) \rightarrow (z = x \vee z = y)))$$

They enable us to make simple numerical statements such as: 'There are five apples on that table' which becomes simply: $(\underset{5}{\exists} x) P(x)$ where $P(x) =_{\text{def}} x$ is an apple on that table.

We can also use the numerical quantifiers to express simple arithmetical equations such as $5+7 = 12$. In fact this becomes:

$$((\underset{5}{\exists} x)\, P(x) \wedge (\underset{7}{\exists} y)\, Q(y) \wedge \neg\, (\exists z)\, (P(z) \wedge Q(z)))$$
$$\rightarrow (\underset{12}{\exists} w)\, (P(w) \vee Q(w))$$

However, as soon as we go on to even quite simple statements of the theory of numbers, the picture changes. Consider, for example, the statement that there is a prime number greater than any given number. If we write: '*m* is a prime number', as '$Pr(m)$', this becomes

$$(\forall n)\,(\exists m)\,(Pr(m) \wedge m > n) \tag{*}$$

However, this involves quantification over numbers which, since they are abstract entities, Kaufmann cannot allow. Moreover, there seems to be no way, using some device such as numerical quantifiers, of reducing (*) to a form in which there is only quantification over material objects. The same applies to the principle of mathematical induction in its first order form:

$$P(0) \wedge (\forall n)\,(P(n) \rightarrow P(n+1)) \rightarrow (\forall n)\,P(n)$$

Here again we have a quantification over numbers which cannot be eliminated.

Thus, if we take the prescriptions of phenomenology seriously, we would be forced to abandon virtually all higher number theory, and most of Peano arithmetic. We would not be able to get much beyond simple numerical equations. This is clearly a very heavy price to pay, and phenomenology is seen to be, in reality, much more restrictive than intuitionism. This argument, incidentally, shows once again that Peano arithmetic goes much beyond what can be obtained by abstracting from the concrete process of counting.

4 KAUFMANN'S DENIAL OF THE ACTUAL INFINITE

As we saw earlier, Kaufmann thinks that we can conceive of the counting process as being capable of being continued without bounds. In this way we get potentially infinite sequences, and we can, if we like, regard such a sequence as denumerably infinite or as having cardinal number $\aleph_0$. It is wrong, however, according to Kaufmann, to think of such a sequence as a completed, actually infinite, totality.

Thus, the lowest transfinite cardinal number $\aleph_0$ is, in a sense, allowable, but all higher transfinite cardinals are regarded by Kaufmann as illegitimate. To get such numbers we need to consider sets of sets, sets of sets of sets, . . ., and such sets are only meaningful if they can be somehow understood as a shorthand for talk of sets of material objects.

In particular Kaufmann rejects Dedekind's definition of real number (p. 104):

> It is thus not possible to obtain irrational numbers as new entities, by means of the procedure of a cut in the domain of rational numbers.

He goes on to remark (pp. 104–5) that Cantor's and Russell's theories of real numbers are equally unacceptable. Indeed for Kaufmann a legitimate real number is simply a sequence of natural numbers generated by some law (p. 102):

> Quite in general, we recognize that an 'infinite decimal fraction' simply means a sequence of natural numbers, where, as repeatedly observed above, we must understand by 'sequence' not an infinite totality, but the domain of a certain relation (law).

If we explicate law here by means of recursive functions, we obtain only denumerably many real numbers. This conclusion would, I think, be acceptable to Kaufmann, for he writes in his summing up (p. 136):

> We can thus not ascend to higher powers than $\aleph_0$. In particular it follows from this that no sense can be attached to the concept of the set of all decimal fractions, of the 'number continuum'. In this way the continuum problem likewise vanishes. Set theory cannot lead us beyond the denumerable; statements that fall into this domain, if correctly formulated, relate not to infinite multiplicities but to formation laws, that is, to arbitrary elements of number sequences.

It might be thought that we could introduce irrational numbers and a non-denumerable real continuum by geometrical intuition instead of set theoretical procedures. Kaufmann does consider this possibility (p. 108):

> In conclusion we may mention the possible objection, that irrational numbers can be intuitively represented by geometrical constructions, let us say $\sqrt{2}$ by the diagonal of a square with side 1, or π by the circumference of a circle with diamter 1.

However, he rejects it on the grounds that geometrical intuition is not sufficiently precise to yield irrational numbers.

Kaufmann's rejection of most of Cantorian set theory provides him with a solution of the contradictions. Thus, regarding the Burali–Forti paradox, he says (p. 158):

> The dissolution of this antimony is effected by realizing that the concept of the set of all ordinal numbers is meaningless.

Kaufmann's approach to set theory would indeed eliminate the contradictions, but unfortunately it would eliminate much of classical analysis as well. However, Kaufmann himself seems unaware of this, and indeed denies that his proposals would result in any significant truncation of classical mathematics (pp. 184–5):

> Once we have seen through this state of affairs, we recognize that the acquisitions of classical mathematics remain untouched by the elimination of the non-denumerable, and the same holds for topology, as far as I can see at present.

5 GENERAL CONCLUSIONS

I could perhaps sum up my general view of Kaufmann by saying that he is strong on the philosophical side, but weak on the technical–mathematical side. Kaufmann gives a good exposition of his basic philosophical principles (his version of phenomenology), and these principles are certainly plausible ones. He goes on to draw out the consequences of this viewpoint within philosophy of mathematics, but here he stops short and does not examine how his critique would affect mathematics. Thus he does not see that his principles would lead to an unacceptable truncation of ordinary mathematics, and instead states, quite erroneously, that they would leave mathematics largely unchanged.

Kaufmann's rather negligent attitude to technical questions is shown in the following quotation (p. 185):

> Since the basic mistake criticized in this essay is deeply rooted in mathematical and logistic semantics, its eradication in many cases encounters enormous difficulties; but these are of a purely technical nature and their solution therefore is in principle of minor philosophic interest, . . .
>
> . . . the practice to which it objects is the ambiguous use of signs, and a refutation of that critique could be seen only in the invalidation of this objection, but not in reference to technical difficulties.

Now of course there is the mistake of losing sight of philosophical issues in a mass of technicalities, but there is also the opposite mistake of failing to examine sufficiently the technical consequences of a particular philosophical proposal. This latter mistake is the one made by Kaufmann.

Generally speaking we may say that there is a tension within philosophy of mathematics between what is mathematically desirable, and what is philosophically plausible. Those (such as Hilbert), who regard existing mathematics as desirable and set out to defend it, find it difficult to discover philosophical principles which are plausible and yet strong enough for the required defence. Conversely those who start out with what seems philosophically plausible either end up with too drastic a truncation of ordinary mathematics (as in the case of Kaufmann), or with the construction of mathematical systems undesirably complicated when compared with the usual ones (as in the case of Brouwer and intuitionism). Kaufmann resolves this basic dichotomy only by ignoring one aspect of it—the mathematical.

Despite this limitation, it still seems to me that Kaufmann's work remains of much value. The problem situation in philosophy of mathematics today, almost fifty years after he wrote, remains much what it was in the late twenties. On the one hand, there is the body of ordinary mathematics, including Cantorian set theory, which is generally accepted, but which really has had no very plausible philosophical justification. On the other hand, there is the constructivist critique. Many mathematicians do feel sympathetic to the constructivist view that much of Cantorian set theory is in some sense 'bogus', 'metaphysical', 'not genuine mathematics'. On the other hand the alternative mathematics offered by the constructivists is too complicated and unworkable to be generally accepted.[1] In this situation it seems natural to look for some middle course between the two principal alternatives. Kaufmann's programme was to supply such a third possibility. Its main features could be put as follows: (1) retention of the law of the excluded middle, and hence elimination of many of the complexities of intuitionism, (2) attempt to make all of mathematics in some sense directly applicable to the material world, and (3) rejection of parts of Cantorian set theory as illegitimate metaphysics.

Kaufmann's attempt to carry out this programme on the basis of phenomenology resulted in an over-drastic truncation of ordinary mathematics; but the programme remains an attractive one, and it might still be possible to carry it out on the basis of a different philosophy. The editor Brian McGuinness is thus greatly to be thanked for rescuing Kaufmann's work from obscurity, and the translator Paul Foulkes for rendering it easily accessible to the English speaking public.

---

[1] Though it should be observed that E. Bishop in his (1967) Foundations of Constructive Analysis, and subsequent authors working in the same general direction, have produced systems of constructive mathematics which are simpler than Brouwer's intuitionism.

D. A. GILLIES
Chelsea College,
University of London

REFERENCES

Bishop, E. [1967]: *Foundations of Constructive Analysis.* McCraw-Hill.

Brouwer, L. E. J. [1975]: *Collected Works,* Vol. 1. North-Holland.

Frege, G. [1884]: *Grundlagen der Arithmetik.* English translation by J. L. Austin [1968], *Foundations of Arithmetic.* Blackwell.

Gödel, K. [1931]: On formally undecidable propositions of *Principia Mathematics* and related systems, I, English translation in J. van Heijenoort (*ed.*) [1967]. *From Frege to Gödel.* Harvard University Press, pp. 596–616.

Hilbert, D. [1935]: *Gesammelte Abhandlungen,* vol. 3. Springer.

Weyl, H. [1918]: *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis.* Veit.

Weyl, H. [1921]: *Über die neue Grundlagenkrise der Mathematik, Mathematische Zeitschrift,* **10,** 39–79.

Brit. J. Phil. Sci. 31 (1980), 299–315 *Printed in Great Britain*

Reviews

DUMMETT, M. [1977]: *Elements of Intuitionism*. Oxford University Press. £9.00. Pp. xii+467.

This excellent and very lucid book aims 'to provide, in a form readily intelligible to one having no prior knowledge of the subject, basic information about the fundamental ideas of intuitionistic mathematics' (p. v). It makes accessible to philosophers trained in mathematical logic ideas and results buried in the technical literature, to which it is intended to serve as an introduction.

Intuitionists take as their fundamental principle (obscure though it is): all mathematical objects, including proofs, are constructed by the mind. If intuitionistic mathematics is what it is supposed to be, it conforms to this principle. The importance of this principle in all aspects of intuitionistic mathematics is a constant theme of Dummett's book, making it not just a text but a philosophical essay on the fundamentals of intuitionistic mathematics as well.

The first three chapters present the basic ideas of intuitionistic mathematics informally. Chapter One explains the notion of proof as construction (constructive proof) and how and why intuitionistic truth and the intuitionistic logical constants are to be understood in terms of this notion. The intuitionistic notion of constructive proof is compared with the notion of constructive proof as it would naturally appear in classical mathematics. The classical view of constructive proof is that such proofs are 'out there' like the natural numbers. Let P be a decidable property of the natural numbers. If not every natural number lacks P, there is a constructive proof that some number has P in the classical sense of constructive proof. (Obtained by checking the natural numbers *seriatim*.) It does not follow that there is a constructive proof in the intuitionistic sense. The intuitionist interprets 'not every number lacks P' to mean that the assumption that every number lacks P leads to absurdity. One might obtain a proof of 'not every number lacks P' in the intuitionist sense of this proposition without thereby having any way to find a number that has P: proofs are not just out there, they must be constructed or constructible according to some known method. Dummett also compares intuitionistic and classical logical principles to show how their differences stem from the intuitionistic notion of proof.

Chapter Two presents the ideas of what Dummett calls 'elementary intuitionistic mathematics': intuitionistic arithmetic, that part of the intuitionistic theory of the real numbers that looks very like the classical theory, order relations among the reals, species (the intuitionistic analogue to sets) and their principal characteristics. The chapter ends with a discussion of axioms of choice and why such axioms are true construed constructivistically.

Chapter Three deals with choice sequences and spreads. Intuitionists conceive of infinite objects as objects constructed without end according to more or less determinate rules for construction. This conception led Brouwer to his theory of choice sequences and spreads, intuitionistic analogues to infinite sequences and collections of infinite sequences. Since to prove a property of a choice sequence is to prove a property of an object not completely constructed, the proof must be based on incomplete information about the choice sequence, on facts which could hold of other sequences. The property holds for any sequence that the incomplete information applies to. The chapter is chiefly devoted to discussing principles and fundamental theorems variously dependent upon this central fact. It also contains a careful examination and elucidation of Brouwer's fallacious proof of the celebrated bar theorem—a proof important for the light it throws on the notion of constructive proof.

Chapters Four through Six deal with the logical analysis of intuitionistic logic and mathematics: formalisations and semantic theories of intuitionistic logic and intuitionistic mathematical theories, and a section on the notion of the creative subject. Dummett deals with all the most important analyses in the literature, save the attempt by G. Kreisel and N. Goodman to provide a direct theory of constructions—he does not do so on the ground that the theory is unfinished. Dummett examines the various analyses with due regard to whether they are faithful to, or help grasp and appreciate, the character and intentions of intuitionistic mathematics. Particularly interesting is a discussion of recent attempts to obtain completeness results for intuitionistic predicate logic by what are called 'generalised Beth trees'. The section on the creative subject deals with attempts to axiomatise the role of the mind in mathematical construction. In some late papers Brouwer introduces the creative subject to define choice sequences that provide counterexamples to certain classical theorems. This introduction has led to much puzzlement. Dummett here provides a subtle contribution to the topic through a discussion on the role of time in intuitionistic mathematics.

The final chapter, entitled 'Concluding philosophical remarks', outlines a defence of intuitionistic mathematics that Dummett has presented in earlier papers (oddly, he does not mention them in the bibliography); discusses the notion of proof; considers whether Beth trees faithfully represent the intuitionistic logical constants; and ends with an account of attempts to axiomatise the notion of choice sequence.

The only criticisms I have worth mentioning are two concerning the first section of the last chapter. First, Dummett does not provide even a cursory exposition of traditional intuitionist philosophical views. If one aims to give basic information about intuitionistic mathematics, one should, I think, explain the views of the chief architect of intuitionistic mathematics, Brouwer. Brouwer's philosophical views are often pooh-poohed but they are no more bizarre than the the views of the early Wittgenstein or most of the views Russell held most of the time. Further, they are in many respects close to the views Dummett puts forward (as the early Carnap is close to the later Carnap). Secondly, the reasoning in this section is surprisingly loose. For example, at the beginning of the section we find: 'If there is a flaw at the heart of classical mathematics, then, even if the intuitionistic reconstruction of mathematics is not correct in every detail, something along those general lines must be right, unless, as is surely unthinkable, all but the most elementary part of mathematics are totally delusory' (p. 360). Either this sentence expresses a non-sequitur or it is an esoteric enthymeme.

There are no important errors, so far as I can tell, except that the proof of the general fan theorem is wrong.

Finally, I want to emphasise that this book is superb, much richer in content than this brief review can indicate.

PETER EGGENBERGER
University of California, Berkeley

SAARINEN, E. (*ed.*) [1979]: *Game-theoretical Semantics* (Essays on Semantics by Hintikka, Carlson, Peacocke, Rantala, and Saarinen). Dordrecht/Boston/London: D. Reidel. Cloth D.fl. 90 (U.S. $39.50). Pp. xiv+383.

It is widely agreed that modern logic owes its spectacular advances first and foremost to the symbolic method invented by Frege and perfected by his followers. By affording formulas which wear their meaning on their sleeves, the logician's ideography enables us to discuss meaning in a rigorous way. In particular, an analysis of a piece of natural language is best given by translating it into an appropriate logical formalism.

At the very time when this method has begun to gain ground among linguisticians, a vigorous group of Finnish logicians have declared war on it. The volume under review is a collection of their opening shots.

We should not, the authors tell us, try to interpret natural-language sentences by translating them into a customary logical notation (p. 170). What we should do instead is to look on such a sentence as an invitation to a *game*. The game in question is for two players, Myself and Nature, and is governed by rules like these:

(*G*.or) When the game has reached a sentence of the form

$$X \text{ or } Y$$

Myself may choose X or Y and the game is continued with respect to it.

(*G*.every) When the game has reached a sentence of the form

$$X\text{—every } Y \text{ who } Z\text{—}W,$$

Nature may choose an individual, give it a name, say b, and the game is continued with respect to the sentence

$$X\text{—}b\text{—}W \text{ if } b \text{ is } a(n)\ Y \text{ and if } bZ.$$

Myself wins if the game ends with a true atomic sentence, and Nature wins if it ends with a false one.

In several places the authors make it sound as if we all spent a great deal of our time actually playing these games, and they invoke the later Wittgenstein as a witness. Be this as it may, it is readily seen that, semantically speaking, little hangs on whether one wins or loses at such a game. When I get beaten at the game associated with a sentence S, it by no means follows that S is false; for I may have lost simply by misplaying my hand. What matters semantically is whether or not there is a foolproof *strategy* whereby I shall win no matter what

Nature gets up to. S is true just in case such a winning strategy exists. And when it comes to establishing that, nothing is gained by actually playing the game.

The game-theoretical definition of truth is undeniably an engaging didactic device and may well hold considerable heuristic potential. But contrary to what the authors would have us believe, it differs from the conventional Tarskian definition in little more than style. One alleged contrast that the authors repeatedly mention and which the present reviewer finds particularly hard to understand is to the effect that the game-theoretical definition proceeds 'from outside in' whereas the Tarskian definition proceeds 'from inside out'. A glance at the recursive clauses making up the Tarskian definition reveals that each of them specifies the semantic value of a compound expression in terms of the semantic values of its immediate constituents, thus taking us, just like a Hintikkian game rule, from outside in. It is true that final semantic verdict on the compound is not reached until the values of the constituents are determined, so that when one eventually hits the rock-bottom atomic sentences one has to retrace one's steps. But parallel retracing surely takes place in the process of determining whether the compound in question admits of a winning strategy, which is the game-theoretical counterpart of truth.

Representing the semantic import of an ordinary-language sentence by means of the associated Hintikkian game has one serious drawback: a representation of this sort is difficult to put on paper. The game associated with a particular sentence S is not defined by simply listing a number of rules like (G.or) and (G.every); for the same rules will normally be used in many different games. The game is rather characterised by the *sequence* in which those rules are applied in various possible plays of the game. Thus to describe the force of S game-theoretically amounts to submitting what Saarinen calls the semantical game *tree* for S (see p. 228). The trouble is that even in quite simple cases, such a tree will be enormous. For instance, if the game involves applications of (G.every) then the tree will feature nodes which split into as many branches as there are individuals in the universe of discourse. It is not surprising, therefore, that in the whole book not a single game is presented in this direct way.

In the case of a perspicuous, canonical language like that of the first-order predicate calculus, semantic game trees can be associated with formulas in a uniform and effective manner by means of a recursive definition. The necessity of actually drawing trees is thus eliminated, for one can write down the corresponding formulas instead. This advantage can be naturally extended to the unregimented sentences of natural discourse. When the tree for such a sentence is inconvenient to draw, why not specify it by way of a perspicuous formula which has the tree in question associated with it by the recursive definition? Such a procedure is, of course, the old translation method over again, which the authors verbally repudiate. But it is exactly this procedure that they in fact apply throughout the book: whenever an analysis of an ordinary-language sentence is offered, it is done by translating it into the notation of first-order logic.

Some of these analyses are eminently interesting. One of them is Hintikka's theory of 'any' and 'every', which in my opinion constitutes a breakthrough in our understanding of these quantifiers. Another is the Hintikka–Saarinen treatment of the Bach–Peters paradox. It represents the most penetrating inquiry so far into a recalcitrant logico-linguistic problem generated by sentences like 'The boy who fooled her kissed the girl who loved him', where two definitely-descriptive phrases involve reference to each other.

Other analyses put forward in the book fail of adequacy and all the failures seem to have a common source: the unwillingness on the part of the authors to venture beyond the idiom of single-sorted first-order logic. Underlying the whole book is an unstated, but consistently adhered to principle whereby the only entities we ever really quantify over or refer to are individuals, and that sets, functions, times, let alone possible worlds never directly come under discussion.

The lengths to which the authors are prepared to go in order to save this principle are sometimes quite extraordinary. The following are three representative examples.

Hintikka detects a hitherto overlooked reading of sentences like 'Some book of every author is referred to in some essay by every critic'. On this reading, the choice of the essay is understood to be independent of the author in question, so that the sentence does not follow from 'The bestselling book by every author is referred to in the obituary essay on him by every critic'. I am leaving aside the question whether English speakers ever understand the former sentence in this way. The relevant point is, as Hintikka shows, that the reading cannot be portrayed in the language of first-order logic. It is, on the other hand, readily portrayed in the language of second-order logic, where we can say that there are *functions* f and g such that for all individuals x and y, if x is an author and y a critic then $f(x)$ is a book by x, $g(y)$ is an essay by y, and $f(x)$ is mentioned in $g(y)$. The obvious conclusion seems to be that on the reading at issue, the sentence makes a second-order statement. Such a conclusion, however, conflicts with Hintikka's preconception that first-order idiom is the universal medium of intelligibility. Hence, he concludes instead that the statement is first-order all right but one in which some quantifiers *branch*, *i.e.*, rather than following one another as is usual, they pile up on top of each other. The first-order principle is thus rescued at the cost of relinquishing the principle of one-dimensionality of logical constructions. To explain the force of the two-dimensional first-order formula offered as a formalisation of the sentence, Hintikka says that it means the same as the corresponding one-dimensional second-order formula. Yet he insists that only the former formula correctly translates the sentence.

Another case in point is Hintikka's treatment of propositional attitudes. Consider the sentence

Tom fancies (the proposition) that everybody admires Tom.

What game rule could possibly be appropriate for sentences of this sort? Since Tom is said to take an attitude to a proposition—*i.e.*, to the set of worlds in which everyone admires Tom—it seems natural to have Myself pick a set of worlds, and endorse two statements: that Tom fancies that set and that it is the set of worlds in which everyone admires Tom. The rule might look like this:

(G.fancies$_p$) When the game has reached a sentence of the form

a fancies (the proposition) that X

Myself may choose a set of worlds, call it p, and the game is continued with respect to:

a fancies p and for all worlds w, w is in p iff X is true in w.

A rule of this sort, however, is at odds with the view that sets of worlds are not to be referred to and worlds are not to be quantified over. Accordingly, Hintikka takes a different approach to propositional attitudes. To deal with 'Tom knows that', for example, Hintikka posits a two-place relation, called the epistemic Tom-alternativeness relation, which a world w bears to all and only to worlds which are compatible with what Tom knows in w. Hintikka's rule then runs like this:

(*G*.knows) When the game has reached world w' and a sentence of the form *a* knows that X Nature may choose an epistemic *a*-alternative w'' to w'. The game is continued with respect to w'' and X.

This approach is objectionable on two counts. First, unlike the rest of Hintikka's game rules, (*G*.knows) makes not just the wisdom, but the very legality of a move depend on matters of fact. In the game associated with 'Tom knows that Bob is a spy', for instance, the first go is Nature's, but according to (*G*.knows), her choice is restricted to worlds compatible with what Tom happens to know. Thus, the semantic game tree for the sentence depends on matters of fact. But the tree is what is supposed to represent the meaning of the sentence. Thus the meaning (semantic import) of the sentence depends on facts. We are saddled with the counterintuitive result that one cannot understand the sentence without knowing what the contents of Tom's knowledge are, in particular without knowing whether Tom knows that Bob is a spy.

The second objection is that the approach is not applicable to propositional attitudes at large. Note that (*G*.knows) relies heavily on the assumption that when one knows something one automatically knows whatever follows from it. Whether or not this assumption is correct, it is undeniable that a parallel assumption would be grossly out of place in the case of other propositional attitudes, like 'fears that', 'is surprised that' *etc*. Tom may fear that Bob is a spy, while being completely resigned to the fact that some people are spies. The fear and surprise relations are similar to the fancy-relation discussed above in that someone may bear them to a proposition without thereby bearing them to its consequences. Tom's propositional fancies are characterisable not by a relation between worlds and worlds, but by a relation between worlds and *sets* of worlds. And the same seems to go for propositional attitudes in general.

The most radical departure from established semantic precepts is advocated in Saarinen's essay entitled 'Backward-looking operators'. Here we are enjoined to relinquish the very principle of semantic functionality for the sake of avoiding quantification over times.

Saarinen's immediate aim is to shed semantic light on sentences such as

(1) Every man who ever supported the Vietnam war now thinks that he was an idiot then.[1]

Those unfamiliar with the latest fashions in logical semantics may be excused for thinking that (1) is not particularly difficult to analyse. Construing the past tense as a two-place relation P between propositions and times (so that $P(p,t)$ means 'p took place at t'), (1) seems naturally formalised thus:

---

[1] The reader is not to be distracted by the extreme factual implausibility of the sentence.

$$(2)\ (\nu)(t)[(\nu \text{ is a man } \& \ P(\nu \text{ supports the Vietnam war}, t)) \\ \supset (\nu \text{ thinks that } P(\nu \text{ is an idiot}, t))].$$

Since (1) contains two past-tense verb forms ('supported' and 'was') it seems natural that (2) contains two occurrences of P. And since 'then' is clearly a temporal pronoun, it is hardly surprising that it should be construed as going proxy for a logical antecedent, *viz.* the implicit time-variable t.

If Saarinen thinks there is something wrong with this obvious way of construing (1), he does not tell us what it is. He seems to take it for granted that the reader shares his intuition that (1) involves no quantification over times and only one application of a past-tense operator, namely, an application which governs everything to the right of 'man'. The reader is thus supposed to be already liberated from the age-old prejudice that in (1) the break after 'man' is logically subordinated to the subject/predicate break after 'war'. What is more, the reader is supposed to find it obvious that 'now' is not an adverb qualifying 'thinks' (and actually expendable in the case in hand), but rather a sentential operator governing the whole of '(he) thinks that he was an idiot then'.

Once (1) is parsed in this peculiar way, the task of accounting for it semantically becomes a formidable one. 'Thinks' is now deemed to be governed by a past-tense operator, yet the thinking is clearly meant to be taking place at the present time. And 'was an idiot' is deemed to be governed by a 'now' operator, yet the idiocy in question is meant to have been contemporaneous with the bygone support for the Vietnam war. Given that no time-variable is present, how does one get from under the 'now'-operator back to the time of the supporting?

Saarinen's parsing method would lead to similar difficulties if applied to the sentence 'Every man who has a mistress tells his wife that he does not know her'. For suppose that someone, taking his cue from Saarinén's analysis of (1), decided not to quantify over mistresses and posited instead two sentential operators: a mistress-operator, governing everything to the right of 'man', and a wife-operator, governing everything to the right of 'mistress'. He would then face the problem of explaining how one gets from the wife back to the right mistress.

But no semantic task is impossible if one is not too choosy about one's tools. Is our problem that of getting from under a 'now' operator back to the time of a previously mentioned event? Why not introduce hitherto unheard-of operators which will do just that? To see how these 'backward-looking' operators work, consider Saarinen's symbolisation of (1):

$$(3)\ (\nu)[\nu \text{ is a man } \supset H(\nu \text{ supports the Vietnam war} \\ \supset N\text{Thinks } D_N(\nu \text{ is an idiot}))].$$

Here H is the always-in-the-past operator, N the now operator, and D_N a backward-looking operator to go with N. The function of the occurrence of D_N is to take you from whatever time is considered at where it stands, to some other time. To determine which time this is, what you do is this: you look at ever larger contexts containing that occurrence until you hit at one which starts with N (in our case it is the clause following the second horseshoe); the time in question is then that corresponding to that larger context.

The idea is simple enough. (Although it would be even simpler and at the same time less restrictive to have, for any number n, an operator D_n which took us exactly n steps back; this would enable us to back up to supercontexts which do not begin with any sentential operator.) What is astonishing, however, is the cheerfulness with which the author brushes aside one of the most fundamental principles of logical semantics without even bothering to mention it. For from times immemorial the task of semantic analysis has been seen as that of explaining the semantic behaviour of an expression in terms of *the semantic behaviour of its own constituents*. What Saarinen has to offer, on the other hand, is an explanation of the semantic behaviour of an expression in terms of *the syntactic features of the contexts in which it is embedded*. It is rather as if someone proposed a novel arithmetic operator f by stipulating:

> an occurrence of '$f(x)$' denotes the sum of x and the first number mentioned to the left of that occurrence.

There is of course nothing to stop someone stipulating this. But by doing so he would surely be leaving the world of arithmetic as we know it. Likewise, there is no stopping Saarinen from travelling through the semantical looking-glass. But the messages he sends back from his journey are bound to meet with the disbelief of those of us who have stayed behind, and indeed evoke Russell's dictum about honest toil and theft.

The reviewer must also confess to being baffled by the general strategy of Saarinen's essay. In the opening sections the problem is stated in informal terms with the help of illustrative examples like (1). Then an artificial language is defined and supplied with a rigorous semantic apparatus. Finally, we are treated to formal paraphrases, such as (3), of the original examples. But on closer scrutiny, the formalisations turn out not to be formulas of the artificial language at all. (Expressions of the type of 'Thinks' in (3) are definitely inadmissible in the language; N seems also inadmissible, although its status is somewhat unclear.) Thus any attempt on the part of the reader to check whether the proposed theory delivers the goods when applied to the author's own examples is frustrated by the bewildering fact that the examples fall outside the purview of the theory.

Limitations of space preclude detailed discussion of the remaining topics dealt with in the volume. They are: conditionals and so-called subgames (Hintikka and Carlson), intensional identity (Saarinen), so-called urn models for first-order logic (Rantala), the Ross paradox (Hintikka), and impossible possible worlds (Hintikka). The volume also contains a critique of Hintikka's treatment of natural-language quantifiers by Christopher Peacocke and Hintikka's reply to it.

With one exception, the book consists of reprints of articles already published in journals and other books. (The exception is Saarinen's essay on backward-looking operators, which, however, will be shortly reprinted in another Reidel volume, also edited by Saarinen himself.) The authors repeatedly apologise for leaving loose ends, the excuse invariably being that there is a limit to what one can fit into an article. Apologies of this sort are bound to sound somewhat lame to the reader who is in fact holding (and may well have paid for) a monograph-sized book in which dozens of pages are wasted on the same points, the same examples, and the same apologies, repeated (often verbatim) over and over again.

PAVEL TICHÝ
University of Otago

WILSON, M. D. [1979]: *Descartes*. London: Routledge and Kegan Paul. Pp. 255. £7·95.

Margaret Wilson tells us in the Preface that her 'interpretation is presented in the form of a sort of semi-commentary on Descartes's *Meditations* Concerning First Philosophy'. She also tells us that she finds the Latin text 'philosophically much more lucid and coherent' than the French. And she tells us that although 'especially strongly influenced' by the work of Frankfurt and Kenny, she has ended up disagreeing with [them] on very many issues of criticism and interpretation'.

A basic difficulty with this 'sort of semi-commentary' is that it pays close heed to the Latin words but little to their historical context. And while she acknowledges the existence of such great French commentators as Alquié, Gouhier, Gueroult and Koyré, none appears to have significantly influenced her reading. Wilson has tried to unravel the central arguments of each Meditation with a minimum of concern either for historical sense or for the interests and methods of twentieth-century philosophy. She wants to avoid losing sight of the Meditations either in history or in our contemporary discussions. By writing a chapter to match each Meditation, she hopes to remind us that she proposes to cast new light on the words of an old text. Like a Reformation Bible commentator, Wilson wants to bring us to the pure scripture without the obfuscations of either the tradition or the modern disputants. Thus, as her choice of form makes clear, the text is the thing. She seeks, wherever possible, to read Descartes sympathetically as well as to call attention to failures in his argument. The result is a solid, measured, reasoned book—but there is no new or daring interpretive thesis.

I have a few quibbles over details. It still seems to me useful to examine the sceptical arguments of the first Meditation in the light of the arguments of Sextus Empiricus and Montaigne. As I think Popkin has established in his [1960], Descartes was coming to grips with a well-formulated and sophisticated philosophical doctrine. The Dream Problem and the quest for criteria are, after all, discussed at some length by the Pyrrhonists. Looking at the issues in that context enables one to distinguish traditional sceptical arguments and their rôles in relation both to so-called sense-experience and to reason. The Deceiver Hypothesis thus marks a new type of challenge and makes the balance of the Meditations an exercise, as has often been noted, in exorcism. As for the Dream Problem, Wilson inclines to the view that Descartes resolves that issue because he 'presumably thinks that our perceptions of the connection and coherence of our experience can be clear and distinct, when we are awake' (p. 29). My impression is that Descartes (and Malebranche after him) did not so resolve the matter, and indeed, saw no satisfactory 'solution' to the Dream Problem beyond strictly restricting ourselves to what is properly clear and distinct; thus, as Wilson emphasises, mathematics need not be affected by the dream question. But despite a weakness on the nature of scepticism, she nevertheless succeeds in striking a just balance between the contending interpretations of the first Meditation and the larger purposes of Descartes' arguments.

In the account of Meditation II, Wilson sets out a number of interpretations of the *cogito* and uses them as devices for exhibiting the plausibility of Descartes' position. The discussion of the piece of wax is again a good example of her ability to cut through the commentaries to the text itself. My only qualms are that she assumes too readily that Descartes accepts a significant version of the primary/secondary quality distinction (p. 80), while her own illuminating account of the ontological status of *mind* and *body* argues against it. Secondly, the otherwise excellent discussion of the essential property of the piece of wax is not tied to the doctrine of innate ideas. Descartes' doctrine of intellectual inspection, and his exclusion of the data of sense or imagination, suggest that innate ideas must be employed to provide an ontological base for extension. That being so, the attempt to impose the language of a primary/secondary quality distinction on Descartes should be pursued cautiously.

My worry about an ontological base for extension extends into Wilson's discussion of Meditation III. I am assuming that Descartes' appeal to innate ideas paralleled the rejection of a relation of likeness or similarity between ideas and properties of material things. The consequence, as Watson in his [1966] has shown, are difficulties in understanding how and in what ways we are to understand the representative character of ideas. Wilson recalls Descartes' distinction between the formal and objective reality of our ideas and goes on to say: 'several texts strongly suggest that when Descartes asks whether an idea represents something real, or *rem*, he is asking whether or not in some way it gives him cognizance of a *possible* existent' (p. 108). Further, 'the claim that an idea or thought is "of a thing (*res*)" must be distinguished *both* from the claim that it represents something that in fact exists, *and* from the claim that it has representational character. An idea may have representational character, yet fail, in the relevant sense, to represent *any thing*' (p. 108). On the one hand, 'considered as modes of thought, sensations have positive formal reality, and like any other real occurrence must have an equally real cause', and on the other, 'sensations, from the objective point of view, are "caused by nothing",' and hence 'sensations must *lack objective reality, despite having representative character*! That they have representative character has been an assumption of the whole discussion. That they lack objective reality follows directly from the premiss that they are "caused by nothing", together with the principle, soon introduced in this Meditation, that there can be no more objective reality in an idea than there is formal reality in its cause' (p. 111).

She wants to hold that 'this divorce between the representative character of an idea and its objective reality is . . . an embarrassment' (p. 111) from the standpoint of proving God's existence from His idea. The difficulty seems to arise from Descartes' account of materially false ideas, *e.g.* ideas of sense. Thus:

'the notion of material falsity provides the basis for an *objection* to Descartes's proof of God's existence, because it entails that the objective reality of an idea is *not* something the idea wears on its face. Descartes would have it otherwise: in his initial exposition of the concept of objective reality he seems to indicate that an idea's objective reality is transparent, deriving directly from its representative character . . .' (p. 112). She goes on to say that I may 'think falsely that my materially false ideas represent something real—that they have objective reality' (p. 112).

Indeed, this is where Wilson wants to take her argument. 'We may wonder . . .' she writes, 'why Descartes should commit himself to the view that materially false ideas *lack* objective reality'. Her answer: 'he was determined at all costs to maintain that the ideas of sense, even if they are *tanquam rerum*, nevertheless fail to exhibit to us any possibly existent quality in an intelligible manner. He may have felt that openly to allow these ideas 'objective reality' would be to undermine his position that in an important ('*de re*') sense they are *not* "of things"' (p. 114). As she concludes (after noting Arnauld's remarks on material falsity): 'The point, again, is that the representative character essential to material falsity necessarily involves a positive content, and must by Descartes's "principles" have a positive cause. In other words, Descartes contradicts himself in saying that materially false ideas come from nothing, from the "objective" point of view' (p. 115).

I have sketched this segment of her argument at some length because I find it perplexing. Although Wilson gives us very little help in understanding what she or Descartes mean by *represent*, the core of her case rests on the tie she claims to find between the 'representative character of an idea and its objective reality'. Reading the Cartesian passages she uses to support her account, I come to a different conclusion. From the objective reality of an idea we can tell, for example, what ontological sort it is an idea of (caused by). It may be the idea of material quality, a substance, *etc*. But of course the objective reality of an idea does *not* carry with it a representing function in the sense of being about an existing thing. Assuming that the latter is what is usually meant by the representative character of an idea, there seems to be no difficulty in arranging the 'divorce'. Sensations do not lack objective reality; in saying they are about nothing and are materially false, Descartes is referring to their failure to represent existent material things (we might be dreaming). But he says on several occasions that their ontological status is grounded in some aspect of our own, rather than another, substance. Ontologically, they are thus not about *nothing*.

Wilson seems to assume that for an idea to be objectively in the mind, it must represent a 'real thing'. Sensations, not representing 'real things' (*i.e.* I may be the cause of their reality), have no objective reality. But her argument seems to go through only if one uses 'represent' ambiguously (caused by an entity in some ontological category *vs*. caused by and similar to some existent) *and* if one then assumes that the objective reality of an idea is its representative function. In brief, the only bridge I can find to her claim that 'materially false ideas *lack* objective reality', is the use of the modern meaning of 'objective'. But I must assume she is not saying that. Accordingly, I again wish that Wilson had employed some historical material in the effort to clarify matters. For example, I wish she had discussed Suarez's formulation (at *Disputationes Metaphysicae* II, § 1) of objective and formal concepts. After all, he is cited by Descartes at that very section of the reply to Arnauld (*HR* II, p. 107) which she finds 'a model of confusion confounded' (p. 110). And she mentions Suarez later in another context.

Finally, a word on her discussion of Cartesian dualism. Although she canvasses a number of Descartes' arguments, Wilson also writes:

> The passage from the *Discourse* [*HR* I, pp. 116–17] provides motivation for Descartes's dualism by suggesting he had reflected—perhaps more systematically than his contemporaries—on the possibility of a mechanistic account of human behavior. It suggests he had concluded that such an account was impossible, in view of the complexity and 'diversity' that would be required in such a machine. If this was his reasoning he must be given credit for, at least, an admirable realism concerning the state of the art and the difficulty of the problem—in contrast, for example, with the unabashed mechanistic optimism of Hobbes (or, as Chomsky might point out, of latter-day behaviorists) (pp. 184–5).

It is refreshing to see Wilson discount various modern interpretations of Cartesian dualism enough to note (I think correctly) that Descartes had commonsensical scientific grounds for his dualism. His defence of dualism was not based solely on those conceptual considerations which alone philosophers now seem to think are worth their attention. Indeed, there is no suggestion that he would have sought to restrain, by the methodological devices favored by many modern empiricists, the application of scientific procedures to mental phenomena.

In conclusion, one of my difficulties with this book is that Wilson apparently feels obliged to take us through sets of alternative analyses of specific problems which she (correctly) considers patently absurd. It is a style which has been mastered in this century only by G. E. Moore and Woody Allen. I have already mentioned my doubts about ahistorical readings of such holy texts. One occasionally gets the sense that Wilson is taking up an issue out of a sense of obligation; more often, one gets the impression that she would like to explore it at much greater length. Nor does the format make it easy to articulate and defend an interpretive thesis. Nevertheless, Wilson has given us a useful book. Precisely because it is a commentary, its parts can be used in relative isolation. Moreover, it is the work of a good philosopher who displays appreciation for and understanding of the intellectual goals of her author.

HARRY M. BRACKEN
McGill University

REFERENCES

Popkin, R. [1960]: *The History of Scepticism from Erasmus to Descartes.* Assen: Van Gorcum.

Watson, J. [1966]: *The Downfall of Cartesianism.*

Earman, J. S., Glymour, C. N. and Stachel, J. J. (*eds.*) [1977]: *Foundations of Space-Time Theories* (Minnesota Studies in the Philosophy of Science, 8). Minneapolis: University of Minnesota Press. Pp. xviii+459. $22.50.

This volume consists of papers arising from two small conferences on the general topic of the foundations of space-time theories held in 1974, one in Minneapolis and one in Andover, the latter held to mark the appearance of the second edition of Adolph Grünbaum's *Philosophical Problems of Space and Time* [1973]. The papers in the book constitute quite a nice microcosm of current work in the philosophy of space and time, and as such portray three important aspects of that work.

The first is simply that the philosophy of space and time *is* now, to a great extent, the foundations of space-time theories. John Stachel sets this out (commenting on earlier sentiments of Euler) in his editor's Notes right at the beginning of the book, and it is reinforced throughout, explicitly by Robert Geroch, by John Earman, and intensely by Howard Stein. The prevailing attitude is that since the nature of space and time are best revealed (perhaps *only* revealed) to us by our best theories, we can do no better than to see what these theories say about them. Of course, this attitude has not always prevailed in just this way; thus, the contribution by Howard Stein (the first of his two in the book), 'Some Philosophical Prehistory of General Relativity', is important because it delves into the relationship between earlier approaches (beginning with Newton and Leibniz) and more modern ones (Riemann and Einstein). There is so much to be said on this subject, and Stein says it so well, that one might wish for more here. And a *bit* more is to be found at the end of Stein's second contribution to the volume. I shall have more to say about that paper in a monent.

Since the natures of space and time as revealed by our best theory—general relativity—are the subject here, it is not surprising that the issues dealt with lie, as John Earman says, 'in a thicket of problems growing out of the intersection of mathematics, physics, and metaphysics' (p. 131). It is the character of the mathematics that seems to me worthy of note as the second aspect of the current work in the philosophy of space and time portrayed in this collection. Prior to the mid-1960s, the overwhelming bulk of work in mathematical relativity theory was carried on by means of 'local' techniques. As Misner, Thorne and Wheeler [1973] have it, 'The non-gravitational laws of physics are obtained by invoking the equivalence principle in a *local* Lorentz frame at each individual event in space-time. To build up an understanding of the global structure of space-time, one performs *local* computations near each event, and then patches the local results together to form a global picture' (p. 916). But in the mid-1960s Roger Penrose at Cambridge University pioneered the use of techniques for analysing global structure directly. Penrose first attacked the problem of the structure of space-time at 'infinity', a region notoriously difficult to analyse locally. These techniques were quickly taken up and applied to the study of other aspects of space-time structure, *e.g.*, horizons, boundaries and singularities, perhaps most notably by Stephen Hawking and Robert Geroch.

This was marvellously creative mathematical physics, full of results of

LIBRARY

striking generality, yet of a very 'picturable' and intuitive character. Best known probably is the result C. J. S. Clarke in this volume calls 'perhaps the most important result in modern relativity theory . . . the prediction by Hawking and Penrose of the necessary occurrence of singularities in general relativistic models of the universe' (p. 99). (This result, and the whole 'global' approach to space-time structure, is best presented in textbook fashion in Hawking and Ellis' [1973] *The Large-Scale Structure of Space-Time.*) But more generally, one finds in this work a host of little theorems and lemmas about the generic properties of curves and surfaces of various kinds in space-times of various structures. These and the characterisations of global causal structure and the techniques for dealing with space-times with various sorts of special regions removed—surgically modified, as it were—have become extremely important in the kind of foundations discussions which appear in this book.

John Earman asks, 'What could it mean to say that time has a beginning or an end? Is it possible that time has a beginning or an end?' (p. 109), and follows the answers straight into the aforementioned thicket. The 'initial bush-beating' Earman undertakes involves subjecting various intuitively reasonable answers to these questions to a variety of examples of space-time manifolds—manifolds, as one might expect, with singularities, boundaries, and other features treated using excisions, truncations, and other exercises of the manifold surgeon's craft. Robert Geroch himself investigates under what conditions one can make predictions and verify them in general relativity, and his analysis, not surprisingly, examines the global causal structure on a variety of surgically modified manifolds. Clark Glymour poses a criterion for the observational indistinguishability of space-time models and discusses quite straightforward examples of such indistinguishability. His analysis involves in a central way the generic properties of certain classes of curves, analysed using global methods. David Malament follows with a companion to Glymour's study, putting the geometrical ideas involved into sharp relief and showing that 'Observational underdeterminism of one sort or another is more the rule than the exception' (p. 62). His paper is simply a tour-de-force of the techniques of global analysis, and presents a striking example of the clarity of statement and depth of analysis of a problem of this sort that can be obtained by their use. The contributions of John Winnie and Lawrence Sklar—in Winnie's case a long and extremely careful study based on a result of Robb concerning the relationship between the causal, topological, and metrical structure of Minkowski space-time, and in Sklar's case an even longer companion piece on the role of axiomitisation in furthering scientific and philosophical purpose—while not depending so strongly on these methods and concepts as do these others, certainly rely on them at important points, as do the papers of C. J. S. Clarke and, in a somewhat different way, Peter Bergmann.

This range of topics encompasses a good deal that is of first importance for the philosophy of space and time. And while these new mathematical methods should by no means be regarded as some kind of philosophical panacea, it indeed appears to be the case that they permit certain traditional issues to be formulated as unprecedently clear questions, and rather well-defined paths staked out toward their answers. As John Earman puts it, 'some of the best philosophy of time is being done today by physicists . . . Conversely, this work inevitably brings a confrontation with traditional philosophical problems' (p. 132).

The third aspect of current work portrayed in this collection of papers is the continuing concentration of work in the philosophy of space and time on the cluster of issues associated with 'conventionality' in this context, issues brought to prominence by the writings of Riemann, Poincaré and Reichenbach. Standing in the centre of this cluster, and in some sense at the centre of this volume is Adolph Grünbaum. The papers of Salmon, Bowman and Friedman all swirl directly around points which Grünbaum has spotlighted, either in his contribution to this volume or in his extensive earlier writings, most notably his *Philosophical Problems of Space and Time* [1973]. The papers by Winnie and by Glymour (evoking the companion papers by Sklar and Malament) while not answering Grünbaum directly, represent work obviously motivated by issues of conventionality which he has emphasised.

A feature that bears mentioning in this debate between the conventionalists—Grünbaum and Salmon in this volume—and their critics—Bowman and Friedman here, less directly Winnie, and, of course, Howard Stein—is that despite the quantity of extremely careful, painstaking (and evangelical and polemical) exposition produced on both sides, very little movement from one camp to the other has taken place. This is certainly not to say that there has been no 'progress' made. Arguments have been refined a good deal, positions have been clarified and extended, helpful metaphors have been developed, and genuinely new results have emerged, while old, somewhat ignored results (*e.g.*, Robb's) have been resurrected, clarified, and brought to bear on the issues.

Still, affiliations have not changed. And distressingly frequently, arguments seem to go *across* each other, rather than *at* each other, as one might hope. That is, almost never does one see a proposition posed that two discussants agree is crucial, and arguments that both see as to the point brought to bear. (One new result in this area might be an exception. As Friedman points out in a note added in proof (p. 431), David Malament [1977] has shown that the standard simultaneity relation can be, in a natural way, uniquely defined in Minkowski space-time. This seems to constitute an argument right to the point on a proposition that all sides seem to count as crucial in the debate about the conventionality of simultaneity.)

This is no better brought out anywhere in print than in the completely engrossing debate presented by Grünbaum's paper here and Howard Stein's reply to it. And this is a real debate, more than a paper and a reply. Grünbaum read a considerably shorter version of the paper at the Andover Conference, where it was actively discussed indeed. Stein subsequently engaged in an extensive correspondence with Grünbaum about this draft, and the paper which appears here is a much expanded version of this Andover draft, with a long section devoted to explicit answers to the points raised by Stein (and others) in correspondence and at the Conference. Stein's piece consists chiefly of one set of his written comments on the original draft, but contains remarks based on Grünbaum's final revised version as well. Thus there are more than two rounds of position and response here—a genuine debate indeed.

And it is a marvellous debate to read—personal and pointed in a way that derives from mutual respect and good will, rich in literary allusion, in metaphor and analogy, sprinkled with rhetorical dash. More substantively, there is historical exegesis of a high order and solid issues of mathematical philosophy, physics, and metaphysics dealt with in a very closely argued way.

Best of all, Stein and Grünbaum disagree here on what seems (despite Stein's contrary avowal) to be just about everything: they disagree resoundingly on the interpretation of Riemann's famous paper 'On the Hypotheses Which Lie at the Foundations of Geometry' [1867] and on the relationship of Grünbaum's position to it; they thus disagree on the relationship of Newton's and Riemann's ideas in general, and here particularly on the interpretation of a letter from Newton to Bentley in 1693 and on *its* import for Grünbaum's position. They disagree on the import of numerous remarks and results of large and small importance made by John Wheeler, J. L. Anderson, John Winnie, Allen Janis, Einstein (one suspects, encouraged by Stachel, p. ix) and, of course, Stein himself. They disagree about the interpretation and import of certain metaphors and analogies, on the turns of literary quotations, and on Leibniz's response in facial expression (at least) to an aspect of Grünbaum's position. Finally, they even differ, says Stein, 'seriously about what to consider a serious difference' (p. 375).

All these differences make the papers extremely valuable, both for an aficionado of the issues involved and for one without this special interest, indeed for someone with little experience in space-time philosophy at all. Putting these papers together with Stein's other contribution to this volume, one obtains as good an introduction to the kind of exegetical issues one must contend with in Newtonian metaphysics and in Riemann's seminal paper as can be found. And Stein's insistence on the most painstaking clarity, his absolutely unyielding determination to employ no word, no concept in systematic discussions whose meaning is not completely unpacked, can hardly be taken too seriously.

This insistence lies at the heart of his criticism of Grünbaum. "What I object to," says Stein, "is that, as it seems to me, you use the terms in question ['ontological' or 'ontologically constitutive', and 'metrical' or 'geometrical'] in a way that *is not clear at all*; or—to state a more modest claim, but with greater assurance—it not only 'seems to me' but is indisputably true that you use those terms in a way that *is not clear to me*" (p. 377). Grünbaum's basic question, as Stein sees it, is 'By virtue of what is the tensor field g to be regarded as representing the metrical structure of space-time?' (p. 377). (Grünbaum formulates it as 'Whence does the metricality of space-time derive ontologically?', p. 327.) Stein immediately reformulates this question to, 'Just what do we *mean* when we say that this tensor field describes the *metrical structure* of space-time?' (p. 375), and proceeds to answer this question *etymologically* by a brief recital of how the phrase 'metric of space-time' came to be employed at all. He is thus left in utter incomprehension of why Grünbaum takes the answer to the question to be one of either physics or ontology.

Of course Grünbaum is at great pains to make just as clear as he possibly can the sense in which he sees issues of physics and ontology to be at stake. Beginning with a sampling of opinions on the 'ontological demarkation' of particles from fields (geometrical or otherwise), Grünbaum embarks on a lengthy treatment of classes of space-times ranging from 'purely empty' to 'non-empty and gravitationally effervescent', arguing for and assessing the implications of, in each case, the claim that "it is the idealized behavior of this important class of external entities [photons, atomic clocks, infinitesimal rods, free massive (gravitationally monopole) particles—his 'family of concordant external metrical standards'] that is taken to be *canonical* for the metricality of these space-times in the sense that it is fundamentally *constitutive* ontologically for their metric structures" (p. 344).

But it is just in this most important section that Stein sees the clearest evidence that Grünbaum simply cannot make himself clear about his fundamental terms, that his position involves a basic incoherence, and thus that the problem of discovering whether the metricality of space-time is ontologically absolute or relational 'may in one way be a quite *impossible* one, because ill-defined—but therefore no problem at all' (p. 374). Stein's analysis is very detailed, and I would say that no one can claim to have fully understood Grünbaum's position if he has not tested his understanding against the barrage Stein throws up.

Toward the end of his comments, Stein further enriches this debate by fastening on what he takes to be a genuine difference of ontological perspective between himself and Grünbaum: "You tend to think of the world in terms of 'things'—'primary substances' in Aristotle's sense. I do not: I tend, rather, to think (Platonically?) of 'structures' and 'aspects of structure' ('Forms'?)" (p. 395). He pursues this difference through a fascinating reading of a difficult passage of Newtonian spatial metaphysics, tying it to a central point from Riemann and hence to his 'prehistorical' remarks at the beginning of the volume: the debate makes a nice package.

Is anyone convinced? John Stachel, in his Notes at the very beginning of the volume, tries to account for 'some of the puzzlement and mutual incomprehension between the two interlocutors' (p. xi), but my guess is that neither Grünbaum nor Stein would accept his account of the basis of their differences: peacemakers always have that problem. My guess is also that the debate will continue, and if it does so in this lively and constructive mode, scholarship in the philosophy of space and time will be the better for it.

ROGER JONES
The University of Tennessee

REFERENCES

GRÜNBAUM, A. [1973]: *Philosophical Problems of Space and Time.* Boston and Dordrecht: Reidel.

HAWKING, S. W. and ELLIS, G. F. R. [1973]: *The Large-Scale Structure of Space-Time.* Cambridge: Cambridge University Press.

MALAMENT, D. [1977]: 'Causal Theories of Time and the Conventionality of Simultaneity', *Nous* 11, pp. 293–300.

MISNER, C. W., THORNE, K. S. and WHEELER, J. A. [1973]: *Gravitation.* San Francisco: W. H. Freeman.

RIEMANN, B. [1867]: *Uber die Hypothesen welche der Geometrie zu Grunde liegen.* (Reprinted in English in M. Spivak: *A Comprehensive Introduction to Differential Geometry, Vol. 2*, 1979, Berkeley: Publish or Perish Press, pp. 135–53.)

CONFERENCE ON SPACE, TIME AND CAUSALITY

Preliminary Notice and Call for Papers

The Royal Institute of Philosophy will be holding a conference at the University of Keele from 4 to 7 September 1981 on 'Space, Time and Causality'. It will deal with philosophical problems about space, time, and causality arising from physics, especially from Relativity Theory and Quantum Theory. The main symposia will be: 'Causality and Quantum Mechanics' (Professor Nancy Cartwright and Mr Jeremy Butterfield), 'Temporal and Causal Asymmetry' (Dr Richard Healey and Mr W. H. Newton-Smith), 'Absolute *vs.* Relative Space and Time' (Mr J. L. Mackie and Mr J. Dorling), 'Time and Causal Connectability' (Professor L. Sklar and Professor Richard Swinburne), and 'Causality, Relativity and the Einstein–Podolsky–Rosen Paradox' (Dr Michael Redhead and Mr Peter Gibbons). The main papers will be circulated in advance of the conference. Booking arrangements for the conference will be announced later.

There will be also an opportunity at the conference for, perhaps, ten contributors to read shorter papers (of about twenty minutes duration) to small groups. Those who would like to read such papers should send an abstract to Professor Richard Swinburne, Department of Philosophy, The University, Keele, Staffs. ST5 5BG to reach him by 1 May 1981.

HISTORY OF SCIENCE

is the only review of literature and research in the history of science, medicine and technology in its intellectual and social context.

FOUNDED IN 1962 AS AN ANNUAL VOLUME, *History of Science* discusses problems and sources for future research in the history of science, medicine and technology, and surveys recent publications. It now appears each year in quarterly parts of about 80 pages each.

VOLUME 18 (1980) will include

- SERGIO MORAVIA *The Enlightenment and Sciences of Man*
- JOHN FORRESTER *Michel Foucault and the History of Psychoanalysis*
- GILLIAN SUTHERLAND AND STEPHEN SHARP *'The Fust Official Psychologist in the Wurrld'*
- JOHN HENDRY *Weimar Culture and Causality*
- C. B. WILDE *Hutchinsonianism, Natural Philosophy and Religious Controversy in Eighteenth Century Britain*
- J. R. JACOB *Restoration Ideologies and the Royal Society*
- ROBERT S. WESTMAN *The Astronomer's Role in the Sixteenth Century*
- J. LONGRIGG *The Great Plague of Athens*

together with other articles, essay reviews and supporting material.

THE ANNUAL SUBSCRIPTION is £16.00 post free outside the Americas and Japan (£11.50 post free to direct private subscribers). New subscribers may purchase the first seventeen volumes (bound) for £85.50 post free, *including* the subscription to Volume 18 (1980). Write to: Circulation Manager, *History of Science*, Whipple Museum, Free School Lane, Cambridge CB2 3RH, England.

THEORIA

A Swedish journal of philosophy
Edited by Peter Gärdenfors, Bengt Hansson
and Krister Segerberg

Volume 44, part 3

Konrad Marc-Wogau Anders Wedberg in memoriam
Philosophical works by Anders Wedberg
Thomas E. Hill, Jr. Kant's anti-moralistic strain
Per Lindström Omitting uncountable types and extensions of elementary logic
John P. Burgess The unreal future

THEORIA, which invites contributions in all fields of analytic philosophy, usually publishes three issues a year. Subscriptions and requests for back numbers should be sent to *Administrative Editor*, THEORIA, *Filosofiska institutionen, Kungshuset i Lundagård, S-223 50 Lund, Sweden.* Subscription for one year is SwCr 55:– or U.S. $ 15.–, for three years SwCr 150:– or U.S. $ 38.–.

Philosophical Studies

An International Journal for Philosophy in the Analytic Tradition

Editors: KEITH LEHRER and JOHN POLLOCK

Philosophical Studies, a journal devoted to work in analytic philosophy, publishes articles in all general areas of philosophical investigation, philosophy of science, epistemology, metaphysics, ethics, and so forth, including the history of philosophy. The main criterion of articles published is that they contain rigorous argumentation and precise analysis of key concepts. Articles in the journal are intelligible to philosophers whose expertise lies outside the subject matter of the article. It is intended that a diligent reader of the journal will be kept informed of the major problems and contributions of contemporary analytic philosophy.

Subscription Information: 1980, volumes 37–38 (eight issues). Private rate: Dfl. 40,– / US $ 20.00 per volume. Institutional rate: Dfl. 124,– / US $ 65.00 per volume. ***Private subscriptions should be sent direct to the publishers.***

D. Reidel Publishing Company

P.O. Box 17, 3300 AA Dordrecht, Holland
160 Old Derby St., Hingham, MA 02043, U.S.A.

Synthese

An International Journal for Epistemology, Methodology and Philosophy of Science

Editors: DAVID GRUENDER and ESA SAARINEN

Senior Advisory Editor: JAAKKO HINTIKKA

Synthese is one of the leading journals in the ford in the philosophy of science and publishes articles in all the fields covered by the subtitle. Special attention is paid to the role of mathematical, logical and linguistic methods in the general methodology of science and, in the foundations of the different sciences, be they physical, biological, behavioural or social. The journal also regularly publishes special issues.

Subscription Information: 1980, volumes 43–45 (nine issues). Private rate: Dfl. 45,– / US $ 22.50 per volume. Institutional rate: Dfl. 135,– / US $ 71.00 per volume. All prices include postage and handling. ***Private subscriptions should be sent direct to the publishers.***

D. Reidel Publishing Company

P.O. Box 17, 3300 AA Dordrecht, Holland
160 Old Derby St., Hinham, MA 02043, U.S.A.

Teaching Philosophy

Contents Volume 3, Number 2

Richard Momeyer	Teaching as a Moral Activity
David Annis/Linda Annis	Does Philosophy Improve Critical Thinking?
Janice Moulton/Jane Voytek	Can Humans Think Machines Think?
Raymond Woller	Beginning with Proofs
Ronald Amundson	Testing Utility
Bangs Tapscott	Natural Deduction and Logical Translation

Special Topic: Teaching Students to Write Philosophy Essays

Peter Spader, Sheldon Richmond, Stephen Daniel, Samuel Gorovitz, Mark Coppenger, and Alastair Gunn

Reviews

Susan Anderson on Philosophy Handbooks, T. M. Reed on Death and Dying, Milton Snoeyenbos and Joseph Hummel on Business Ethics, plus Kurt Baier, Robert Brumbaugh, Patricia Johnston, Campbell Crockett, and others.

Subscriptions (4 issues): $12 individuals; $20 others. Add $2 for postage outside USA. Send orders to the Philosophy Documentation Center, Bowling Green, OH 43403.

Edited by Arnold Wilson, University College, University of Cincinnati, Cincinnati, Ohio, 45221.

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY

Edited by MARVIN FARBER

JUNE 1980, VOLUME XXXX, No. 4

CONTENTS

I. ARTICLES

Intentional and Mechanistic Explanation, by VIVIAN M. WEIL
Transformations, by WILLIAM SACKSTEDER
A Dialogue on Selfishness, by A. MacC. ARMSTRONG
Quid Facti or Quid Juris: The Fundamental Ambiguity of Gadamer's Understanding of Hermeneutics, by LAWRENCE M. HINMAN
Conscientious Disobedience of the Law: Its Necessity, Justification, and the Problems to Which it Gives Rise, by H. J. McCLOSKEY

II. DISCUSSIONS

III. REVIEWS

IV. RECENT PUBLICATIONS

Subscription prices $10.00 per year for individuals; $12.00 per year for institutions. Separate issues, $2.50 and $3.00, respectively. Postage and handling for back issues only. Papers submitted for publication will not be returned unless accompanied by a self-addressed, stamped envelope, or equivalent postage.

A Circular listing the main contents of the Journal since 1940 will be sent upon request.

Address all communications to

PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH
State University of New York at Buffalo,
609 Baldy Hall, Buffalo, New York 14260

Journal of Philosophical Logic

Published under the auspices of the Association for Symbolic Logic

Editor-in-Chief

RICHMOND H. THOMASON
University of Pittsburgh, U.S.A.

The Journal of Philosophical Logic is the only journal specializing in philosophical logic and utilizes formal methods of dealing with topics in logical theory, including contributions to branches of logical theory directly related to philosophical concerns, contributions to philosophical discussions that utilize the machinery of formal logic, discussions of philosophical issues relating to logic and the logical structure of language, and philosophical work relating to special sciences with emphasis on foundational problems.

Subscription Information

1980, vol. 9 (four issues)
Private rate: Dfl. 45,– / US $ 22.50/ *Institutional rate*: Dfl. 139,– / US $ 73.00
Private subscriptions should be sent direct to the publishers

D. Reidel Publishing Company
P.O. Box 17, 3300 AA Dordrecht, Holland
160 Old Derby St., Hingham, MA 02043, U.S.A.

The Monist

An International Quarterly Journal
of General Philosophical Inquiry

Founded 1888 by **Edward C. Hegeler**

Editor, Eugene FREEMAN

Editorial Board: William P. Alston, Monroe C. Beardsley, Lewis White Beck, William A. Earle, Dagfinn Føllesdal William Frankena, Maurice Mandelbaum, R. Barcan Marcus, Richard Martin, Mary Mothersill, Joseph Owens, Richard Rorty, J. B. Schneewind, Wilfrid Sellars, John E. Smith, Richard Wasserstrom.

Managing Editor: Sherwood J. B. SUGDEN

EACH ISSUE IS LIMITED TO ARTICLES ON A SINGLE GENERAL TOPIC. Communicate in advance with the Editor for special instructions. Two nonreturnable copies (one copy outside the U.S. and Canada) should be submitted NINE MONTHS PRIOR to the announced publication date. Papers should be prepared for anonymous submission to readers, and should be from 10 to 20 double-spaced typewritten pages in length, including notes.

SCHEDULED PUBLICATION DATES:

Vol. 64, No. 4	Oct.,	1981	Heidegger and the History of Philosophy
Vol. 65, No. 1	Jan.,	1982	Contemporary Continental Analytic Philosophy
Vol. 65, No. 2	Apr.,	1982	The Carus Lectures of Donald Davidson (By Invitation Only)
Vol. 65, No. 3	July,	1982	The Philosophy of Wilfrid Sellars
Vol. 65, No. 4	Oct.,	1982	Metaphysics and Logic
Vol. 66, No. 1	Jan.,	1983	Right Reason in Western Ethics
Vol. 66, No. 2	Apr.,	1983	Integrity in Art and Morals
Vol. 66, No. 3	July,	1983	Categories

Editorial Office: Box 1908, Los Gatos, California 95030
Business Office: Box 600, La Salle, Illinois 61301

SUBSCRIPTION RATES: United States: Annual (4 issues): Institutions, $18.00; Individuals, $12.00. Single copies: $4.50. Foreign postage: Add 25 cents to single copy rate or $1.00 to subscription.

AUSTRALASIAN JOURNAL OF PHILOSOPHY

EDITOR: BRIAN ELLIS

CONTENTS OF VOLUME 58, NO. 1, MARCH 1980

ARTICLES	
Substance Substantiated	C.B. Martin
Believing in Sentences	John Bigelow
McGinn and Essential Properties of Natural Kinds	Laurance J. Splitter
A Note on Physicalism and Heat	Frank Jackson
Functionalism and Phenomenalism: A Critical Note	Colin McGinn
Expressions of Ignorance	B.H. Slater
DISCUSSION	
Privacy and Respect for Persons: A Reply	S.I. Benn
Jackson on Counterfactuals	Wayne A. Davis
CRITICAL NOTICE	
Rational Belief Systems by Brian Ellis	Brian Skyrms

REVIEWS, NOTES AND NEWS

ANNUAL SUBSCRIPTIONS: Individuals A$13.00 (US$15.00 or £stg.7.50); Institutions A$20.00 (US$22.00 or £stg.11.00). All correspondence should be addressed to the Editor or the Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY
by
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

STUDIA LOGICA

An international quartely for symbolic logic and its applications in formal philosophy of science

Editor in Chief: Ryszard WÓJCICKI, P.O.Box 79, 90-954 ŁÓDŹ 4, POLAND

Volume 38, Number 3 (1979)

Contents

George F. Schumm, Modal logics with no minimal proper extensions

Roberto Cignoli, Coproducts in the categories of Kleene and three-valued Łukasiewicz algebras

Tadeusz Prucnal, On two problems of Harvey Friedman

Piotr Wojtylak, Matrix representations for structural strengthenings of a propositional logic

Anita Wasilewska, A constructive proof of Craig's interpolation lemma for m-valued logic

Arnold vander Nat, First-order indefinite and uniform neighborhood semantics

Aldo Ursini, A modal calculus analogous to K4W, based on intuitionistic propositional logic, I°

Published by Ossolineum Publishing House and North-Holland Publishing Company. Subscription orders: from the socialist countries — RSW „Prasa-Książka-Ruch", Wronia 23, 00-840 Warszawa, Poland; from outside the socialist countries — North-Holland Publ. Co., P.O. Box 211, 1000 AE, Amsterdam.

SCIENTIA

International Review of Scientific Synthesis - Founded in 1907

Scientific Editorial Board: Piero Caldirola · Ludovico Geymonat · Giuseppe Montalenti
Editor: N. Bonetti

SCIENTIA in 74 years has never abandoned its programme of rigorous scientific synthesis; it aims at going beyond rather than refuting disciplinary specialization.

The motivations behind this approach can be explained today on the one hand through the formation of new areas of scientific research, and on the other through the lively evolution of those social frameworks which are at one and the same time both cause and effect of such research.

Scientia takes on the task it has set itself with the help of contributions from scientists and philosophers from all parts of the world.

Publication: 3 volumes a year of approx. 400 pages each. Articles are published in the language in which they were written and in the integral English and Italian versions.

Annual subscription: Italy Lit. 22.000 - Europe Lit. 27.000 - Non-European countries $44.00
Specimen copy. Italy Lit. 6 000 - Europe Lit. 5.500 - Non-European countries $ 7.50

SCIENTIA - Via Guastalla 9 - 20122 Milano (Italy) - Telephone (02) 780.669

INQUIRY

Editor: Alastair Hannay
Founding and Consulting Editor: Arne Naess

Volume 22 (1979), No. 4

Are Luddites Confused?	Dan Lyons
The Centrality of Problem-Solving	John Kekes
The Roots of Epistemological 'Anarchy'	Nancy J. Nersessian
Sherlock Holmes – Philosopher Detective	Wulf Rehder
Making Sense of History: Skagestad on Popper and Collingwood	M. Hurup Nielsen and Jeremy Shearmur

Volume 23 (1980), No. 1

Democracy, Elitism, and Scientific Method	Paul Feyerabend
Science, Reason, Knowledge, and Wisdom: A Critique of Specialism	Nicholas Maxwell
Creative Product and Creative Process in Science and Art	Larry Briskman
Epistemological Relativism in its Latest Form	Harvey Siegel
Siegel on Kuhnian Relativism	Gerald Doppelt
A New Science of Criticism	Price Charlson

Subscription:

Foreign: Institutions: N.kr 130,–, $26.00 – Individuals N.kr 78,–, $16.00.
Norway: Institutions: N.kr 98,–, Individuals: N.kr 58,–.

Published quarterly by Universitetsforlaget, P.O. Box 2959 Tøyen, Oslo 6, Norway – P.O. Box 258, Irvington-on-Hudson, New York 10533, USA.

the review of
metaphysics

a philosophical quarterly

JUNE 1980 | VOL. XXXIII, No. 4 | ISSUE No. 132 | $5.00

articles

KENNETH L. SCHMITZ	*A Moment of Truth*
EDWARD POLS	*The Ontology of the Rational Agent*
ROBERT S. BRUMBAUGH	*Time Passes: Platonic Variations*
JIM K. SWINDLER	*Parmenides Paradox: Negative References and Negative Existentials*

critical study

RICHARD J. BERNSTEIN	*Philosophy in the Conversation of Mankind*

books received

DEREK CROSS and Staff	Summaries and Comments

philosophical abstracts *announcements*

Individual Subscriptions $15.00 Institutional Subscription $20.00 Student Subscriptions $9.00

Jude P. Dougherty, Editor
The Catholic University of America, Washington, D.C. 20064

Zeitschrift für allgemeine Wissenschaftstheorie

Journal for General Philosophy of Science

Edited by ALWIN DIEMER . LUTZ GELDSETZER . GERT KÖNIG

Per year 2 issues of 208 pp. each Annual subscription DM 92,–

From the contents of vol. X/1 1979

AUFSÄTZE

LOUIS BOON, Repeated Tests and Repeated Testing: How to corroborate low level hypotheses

KLAUS JÜRGEN DÜSBERG, Sind empirische Theorien falsifizierbar?

ULRICH HOYER, Das Verhältnis der Leibnizschen zur Keplerschen Himmelsmechanik

JOSEF MERAN, Individualismus oder Kollektivismus? Versuch einer Rekonstruktion eines sozialwissenschaftlichen Grundlagenstreits

TORE NORDENSTAM–HÅKAN TÖRNEBOHM, Research, Ethics and Development

GERARD RADNITZKY, Das Problem der Theorienbewertung. Begründungsphilosophischer, skeptischer und fallibilistischer Denkstil in der Wissenschaftstheorie

ROLF SCHOCK, On Classifications and Hierarchies

EFRAIM SHMUELI, How is Objectivity in the Social Sciences Possible? A Re-evaluation of Karl Mannheim's Concept of "Relationism"

DISKUSSION

HAROLD I. BROWN, A Functional Analysis of Scientific Theories

NICO STEHR–ANTHONY SIMMONS, The Diversity of Modes of Discourse and the Development of Sociological Knowledge

BERICHT

LUTZ DANNEBERG–HANS-HARALD MÜLLER, Verwissenschaftlichung der Literaturwissenschaft. Ansprüche, Strategien, Resultate

---

FRANZ STEINER VERLAG GMBH . D-6200 WIESBADEN
W. GERMANY P.O.B. 5529

EDITORIAL BOARD

R. B. Braithwaite
G. Buchdahl
L. J. Cohen
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
W. C. Kneale
S. Körner
J. L. Mackie
D. J. O'Connor
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow
J. H. Woodger

EDITORIAL ADVISORY PANEL

The Editors are helped by members of the Editorial Board, members of the Committee of the British Society for the Philosophy of Science and by an Advisory Panel which includes:

Peter Clark
Jon Dorling
Arthur Fine
Michael Gardner
Donald Gillies
Ian Hacking
Colin Howson
Clive Kilmister
Noretta Koertge
David Miller
Alan Musgrave
Vernon Pratt
Ian Roxburgh
Geoffrey Sampson
Peter Urbach
Elie Zahar

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1980-81

President: Mr G. Buchdahl
President-Elect: Professor C. W. Kilmister
Hon. Secretary: Mr G. Hunt
Hon. Treasurer: Mr J. Powers
Members: Dr D. A. Gillies, Dr C. Phillips, Dr M. L. G. Redhead, Professor Y. Wilks

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all interested persons on payment of the annual subscription. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £2.50 for U.K. residents and £1.25 for overseas residents ($2.50 if resident in U.S.A. or Canada). Those who wish to join should write to the Hon. Secretary, Mr G. Hunt, Department of Philosophy, University of Warwick, Coventry CV4 7AL.

Brit. J. Phil. Sci. Vol. 31 No. 3, September 1980

The British Journal for the Philosophy of Science

Contents

Articles

213–254 John P. Mayberry A New *Begriffsschrift* (I)

255–272 Richard Swinburne Conventionalism About Space and Time

Discussions

273–282 Paul Feyerabend Zahar on Mach, Einstein and Modern Science

282–288 Hugh Mellor On Things and Causes in Spacetime

Review Article

289–298 Donald Gillies Phenomenology and the Infinite in Mathematics (review of F. Kaufmann: *The Infinite in Mathematics*)

Reviews

299–315 Peter Eggenberger on Michael Dummett: *Elements of Intuitionism*; Pavel Tichý on Esa Saarinen (*ed.*): *Game-theoretical Semantics*; Harry Bracken on M. D. Wilson: *Descartes*; Roger Jones on J. S. Earman, C. N. Glymour and J. J. Stachel (*eds.*): *Foundations of Space-Time Theories.*

Announcements

17 JAN 1981

Printed in Great Britain at Aberdeen University Press

Volume 31 Number 4 December 1980
BJPIA5 31 (No. 4) 317-416 (1980) ISSN 0007-0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: JOHN WORRALL

Subscriptions

The *Journal* is published quarterly. The subscription price for mailing addresses in Britain is £7.00 net per volume post free and for mailing addresses outside Britain £11.00 (or foreign currency equivalent calculated at the day's rate); single issues £2.50 net (overseas £4.00) plus postage. Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

For members of the British Society for the Philosophy of Science—who should order direct from Aberdeen University Press—the subscription price is £4.50 (overseas £7.00). Membership of the Society is international and is by no means confined to people resident in the United Kingdom.

Microform

The current volume of the *Journal* is available in microform to subscribers to the printed version, *at a combined print/microfiche rate of* £12.50 *(overseas* £16.50).

Back Volumes

Volumes 1–16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17–30 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application. Volumes 1–29 are also available in microform from Kraus Reprint Corporation.

Typescripts and Review Books

Typescripts, correspondence and books for review should be sent to the Editor at the Department of Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street, London WC2A 2AE.

Review books are accepted on the understanding that, although they will appear in the 'Recent Publications' list, a full review cannot be guaranteed, and that books not reviewed are not returnable.

Notes to Contributors

Three copies of all contributions should be sent, typed on one side only with wide margins and double-spaced. Detailed instructions on the lay-out of typescripts are published occasionally at the end of the *Journal* and anyway may be obtained from the Editor.

Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

© British Society for the Philosophy of Science 1980

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editor.

Brit. J. Phil. Sci. 31 (1980), 317–328 *Printed in Great Britain*

Social Propensities

by PETER URBACH

1 *Introduction*
2 *Physical and Social Propensities*
  (*a*) *Physical Propensities*
  (*b*) *Social Propensities*
3 *Social Propensities Vindicated*
4 *Statistical Regularities and Holism*
5 *Conclusion*

1 INTRODUCTION

In his famous [1897] study of suicide, Emile Durkheim made various striking discoveries. Most notably he found that the annual rate of suicide in each of the social groups studied was extremely stable over many years and this stability far exceeded that of the general mortality rate. Durkheim also noticed a striking and persistent variation in the suicide rate from country to country. Thus, while Denmark's suicide rate was on average 277 per million per year, the corresponding rates for Austria and Italy were just 78 and 30 respectively. Also, within each country different social groups exhibited a regular order in their susceptibility to suicide. For example, within any country, the suicide rates followed the order:

$$\text{Protestants} > \text{Jews} > \text{Catholics}.$$

Some doubt has been thrown on Durkheim's figures in recent years but nothing of what follows depends on their accuracy since I shall be concerned with methods of explanation in social science rather than with the merits of any particular explanation. Durkheim used his data on suicide rates to establish the need for holistic explanations in the social sciences. He discussed the possibility of providing an individualistic explanation of the suicide characteristics in each society but found that no plausible candidate fitted the facts and the striking conclusion he then drew was that:

> The social suicide rate can be explained only sociologically. At any given moment the moral constitution of society establishes the contingent of voluntary deaths. There is therefore for each people a collective force of a definite amount of energy, impelling men to self-destruction. The victim's acts, which at first

Received 30 April 1979

> seem to express only his personal temperament, are really the supplement and prolongation of a social condition which they express externally. . . . It is not a mere metaphor to say of each human society that it has a greater or lesser aptitude for suicide; the expression is based on the nature of things. Each social group really has a collective inclination for the act, quite its own. . . . The private experiences usually thought to be the proximate causes of suicide ***have only the influence borrowed from the victim's moral predisposition, itself an echo of the moral state of society*** (*ibid.*, pp. 299–300; my italics).

> Usually when collective tendencies or passions are spoken of, we tend to regard these expressions as mere metaphors and manners of speech with no real significance but a sort of average among a certain number of individual states. They are not considered as things, forces *sui generis* which dominate the consciousness of single individuals. Nonetheless this is their nature, as is brilliantly shown by the statistics of suicide (*ibid.*, p. 307).

> Collective tendencies have an existence of their own; they are forces as real as cosmic forces (*ibid.*, p. 309).

The cause of holistic social explanations could hardly be more clearly and unequivocally avowed and it is just this sort of explanation which Popper, Watkins, Hayek and others had in mind in their well-known attacks on holistic social science.[1]

Watkins characterises methodological individualism as the principle

> that social processes and events should be explained by being deduced from
> (*a*) principles governing the behaviour of participating individuals and
> (*b*) descriptions of their situations,

while methodological holism states that the behaviour of individuals should be explained by being deduced from

> (*a*) macroscopic laws which are *sui generis* and which apply to the social system as a whole, and
> (*b*) descriptions of the positions (or functions) of the individuals within the whole (Watkins [1952], p. 150).

The essential distinction between individualistic and holistic explanation is, perhaps, not entirely clear from Watkins's characterisation. It is clear, however, that an important part of individualism is that the crucial and irreducible causal factors in social change are *mental* processes, such as beliefs and preferences, while holism postulates *collective forces* which are properties of the community as a whole and which may produce socially significant changes.

Thus in order to explain why a war erupted between two countries the individualist might describe the feelings of dislike, fear *etc.* of the relevant people, their beliefs and decision rules. And with the assistance of the so-called 'rationality principle', the individualist will hope to deduce a description of a state of affairs which can be, with the help of some definitions, characterised as a state of war.

---

[1] For an excellent collection of relevant articles, see O'Neill [1973]. Page references to Watkins's articles are made to this work.

On the other hand, a holistic explanation for the same war might, conceivably, take the form of a statement that all societies of type W go to war every n years, S is such a society whose previous war occurred $n+1$ years ago, hence this society went to war last year. This is, of course, a rather unrealistic example. Herbert Simon's holistic explanation for the approximate linear relation between the population size of American cities and their rank order in population size is rather more believable. The explanation for this remarkable phenomenon includes the assumptions that the 'probability that any migrant, chosen at random, will migrate [from a rural area] to a city of a particular size class is proportional to the total urban population in that class of cities', that the emigration rates from cities are uncorrelated with city size, and that 'of the total growth of population in cities above some specified minimum size, a constant fraction is contributed by the appearance of new cities [of that size]' (Simon [1977]).

It is sometimes suggested that holistic explanations are bound to be *ad hoc*, but the two holistic explanations which I have adumbrated are certainly not that; although, perhaps, they have other defects.

2 PHYSICAL AND SOCIAL PROPENSITIES

Durkheim frequently speaks of the collective 'propensity' of a people to suicide as a synonym for the '*sui generis* forces' which he postulates as the causes of the suicide rates. Popper ([1957] and [1959]) used the term 'physical propensities' to characterise certain stochastic processes in purely physical settings and in this section I will discuss the significant similarities between physical and social propensities.

(*a*) *Physical Propensities*

The probability calculus is defined in a purely formal way in terms of axioms and rules of inference. Thus, an axiomatisation of the calculus may contain formulas such as

$$p(a, a) = 1$$

$$0 \leqslant p(a, b) \leqslant 1$$

$$p(a \,\&\, b, c) = p(a, c)p(b, a \,\&\, c)$$

Despite the almost universal agreement over the formal laws of the calculus, there is still a great deal of discussion over the question of the most appropriate interpretation of the various symbols appearing in it.

One of the most widely used interpretations associates infinite event classes with the functional variables *a*, *b*, *c* . . . and gives a meaning to the probability functor, *p* in terms of the relative frequency of events. Thus according to the frequency interpretation, the term $p(a, b)$ means the limiting relative frequency of type *a* events in a random sequence of type *b* events.

Popper's criticism of the frequency approach and his reasons for replacing it with his propensity interpretation can be readily summarised. He considers a die which is loaded so that in long runs of throws it yields a relative frequency of showing a six uppermost of 1/4; he also considers a sequence consisting of throws of the loaded die together with just three throws of a homogeneous symmetrical die, which is then immediately destroyed (let us say). Clearly one would be inclined to think that the probability of throwing a six was different for each of the dice, in particular that the probabilities of a six appearing was 1/4 and 1/6 for the loaded and fair die, respectively. However, the frequency theory seems to exclude this belief since the fair die was only thrown three times and could not possibly have yielded a relative frequency of sixes of 1/6.

In order to be able to retain the belief that the fair die has a probability of yielding sixes of 1/6, the frequentist would usually say that had this die been thrown a large number of times (which it of course wasn't) then the relative frequency of sixes in this virtual sequence would have tended to 1/6. But in referring to a virtual sequence one must refer to a set of generating conditions since the sequence must be one produced by the repetition of the same experiment. But, Popper now says, if the frequentist makes this modification then the characteristic feature of frequentism falls away; probability is now not the property of a sequence which is given but the property of a set of a sequence-generating conditions, that is of some physical set up. This is precisely the interpretation which Popper argues should be explicitly adopted; and according to this propensity interpretation the experimental conditions are endowed with a tendency, or disposition or propensity to produce sequences whose long run relative frequencies are numerically equal to the probabilities. This is the essence of Popper's argument, and a very good argument it is.

Physical propensities are thus dispositional properties of an experimental arrangement and they 'resemble forces, or fields of forces'. And, like Newtonian fields of forces in a physical system, propensities are 'always a property of the *whole physical system*' (Popper [1959], pp. 37–8).

The determined positivist might, perhaps, reject abstract dispositional properties as unreal on the grounds that, unlike red, say, they are not immediately perceivable; but this kind of objection if thoroughly pursued would also dispose of the whole gamut of useful scientific notions from 'atoms' to 'gravity' and it would imply that a die which had only ever been thrown a few times just did not have any probability properties at all.

We can, I think conclude with Popper that 'propensities are not only as objective as the experimental arrangements but also *physically real*—in the sense in which forces, and fields of forces are *physically real*' (Popper [1957], p. 69).

(*b*) *Social Propensities*

Now it seems to me that Durkheim's notion of a collective social propensity to suicide does not merely reflect a verbal similarity to Popper's theory of physical propensities but rather that the two are essentially analogous.

We will consider an 'experimental set up' which consists of selecting persons at random from a social group and observing whether they commit suicide or not within some time period. Durkheim's analysis of the results of such 'experiments' showed that within each social group long sequences of outcomes yield increasingly stable relative frequencies of suicides and that each social group had its own characteristic frequency of self-destruction. Moreover, there appeared to be no means for predicting in any case whether an individual would commit suicide or not or exactly what proportion of the population kill themselves in each year.

Just as scientists conjecture that there is a definite probability of ones and sixes when a die yields increasingly stable relative frequencies in apparently random sequences of throws, Durkheim surmised the existence of a definite probability for suicide in each social group.

At present we have in mind a purely frequentist notion of probability, so that Durkheim's conjecture amounts, roughly, to nothing more than the claim that if the sequences of outcomes in each 'experiment' are lengthened indefinitely then the relative frequencies of suicide will tend ever closer to the probability value.

A question now arises. Why do distinct probabilities attach to different countries and social groups? In the case of Popper's dice, the major influences were their symmetry and mass distributions. Durkheim identifies the degree of 'social integration' in a group as the crucial determinant of the probability of suicide. (Let us assume—counterfactually, I think—that there is a satisfactory independent measure of social integration.)

Now, continuing the parallel with Popper's argument, suppose that three Danes emigrate to Italy, a country whose suicide rate is much below that prevailing in their homeland. Might one not now wish to attach to these immigrants a probability of suicide different to that of the indiginous population? Or suppose a small Protestant sect were to spring up in a mainly Catholic milieu. We may know that the level of social integration is much diminished by the religious conversion and our theory would predict a significantly increased probability of suicide. But now suppose that this new Protestant sect is unfortunate enough to have settled in a region of acute seismic disturbances and that it is in fact wiped out by an earthquake after, say, three month's existence. As with the fair die which was destroyed after only three throws, there can be no probability of suicide in the frequentist sense for the Protestant sect, unless one is inclined to consider virtual sequences and would-be suicides. But then, mimicking Popper's argument, when we characterise the virtual sequence we must refer to the conditions generating that sequence, that is to say, we must specify the relevant social arrangements in the group and this is tantamount to associating a dispositional property or propensity to the social set up as a whole.

It seems then that by an excellent argument supplied by Popper, Durkheim violated no canon of science by postulating social propensities which are 'forces as real as cosmic forces'. Of course, it may turn out that these social propensities are reducible to more mundane things so that we might one day be in a position to specify conditions which determine precisely who commits suicide in any given year. The same obviously applies to physical propensities. We are, however, not discussing the truth of propensity theories, a matter which could not be decided anyway, but whether such theories are scientifically respectable in the social sciences.

3 SOCIAL PROPENSITIES VINDICATED

Durkheim's own attempt to produce a satisfactory holistic propensity theory of suicide was probably *ad hoc*. However, the postulation of social propensities does not always lead to *ad hoc* explanatory theories and in fact modern theories do give evidence for the existence of such propensities. I have in mind the reproductive system as a propensity set up.

We each possess, according to the established theories of biology, twenty-three pairs of chromosomes along which hereditary units known as 'genes' are ranged. Each gene on a chromosome has a corresponding gene, or 'allele', on the associated chromosome of the pair. Now some of the genes are of social importance; their assortment in a population determines, at least to some degree, the distribution of physical and psychological characters within the group. Some of these inherited characters, especially emotional and intellectual ones, are of profound social significance. and even a slight change in their distribution in a population may produce quite distinct social formations.

According to modern Mendelism the genetic constitution of an individual is ***not uniquely determined*** by that of his parents. Instead genetic theory attaches weights or probabilities to the various possible genetic constitutions. For instance, two individuals each with the genes A and a at corresponding loci on a chromosome may yield offspring bearing the gene pairs: AA, Aa and aa. Genetic theory predicts that these gene pairs will occur with a probability of 1/4, 1/2, and 1/4, respectively. The probabilities of the possible distributions of these genetic pairings in the over-all populations depend on several factors, including the genetic constitution of the parent population and its mating patterns and the latter will clearly depend on the customs beliefs and institutions of the society. Meritocratic societies for instance will, other things being equal, mate assortatively for intelligence. A society composed of isolated settlements will lead to a high rate of inbreeding, which could be reduced say, by the introduction of a railway system. If one section of the population, like the priesthood, did not produce children, the probability of transmission from generation to generation of inherited sacerdotal characters (if there are any) will diminish. There can be little doubt that the cultural character of a society will influence which genetic types produce children, what kinds of mating patterns prevail, and how many children result from different matings. It is thus that the cultural character of a society influences the propensity of that society to produce certain types of individuals. We may easily conceive how propensities akin to those which Durkheim had in view could be produced and why they would vary between different societies. Suppose an inherited and measurable psychological character is normally distributed and that its possession to more than degree k determines an individual to take his own life. The number of individuals within a society who are afflicted by the suicidal character is, as I said, not uniquely determined by the reproductive system. But each society will be characterised by a propensity to produce different numbers of suicides and this propensity will be determined by the society 'as a whole', in the sense that its value depends partly on the prevailing cultural values, beliefs and institutional arrangements.[1]

So far we have seen that if irreducible physical propensities are scientifically acceptable in natural science then irreducible social propensities are also acceptable components of theories about societies. And just as physical propensities are 'relational properties of an experimental arrangement' and 'a property of the whole physical system' (Popper [1959], pp. 37–8). so social propensities of the types discussed here are properties of the whole social system.

---

[1] Other factors such as physical conditions and diet will, no doubt, also play a part.

I have up till now emphasised social propensities arising out of the reproductive system since these are based on well-confirmed scientific theories and their existence can thus be plausibly supposed. But there is no reason why other propensity set ups of a non-genetic type should not exist. These propensities could be more like those Durkheim probably envisaged in which the social structure exerts an inclining, though not compelling, force on individuals who already find themselves in that structure, rather than the propensities I have argued exist in virtue of the reproductive system and which are tendencies to produce given types of individual.

4 STATISTICAL REGULARITIES AND HOLISM

The significance for individualism of statistical regularities has been commented on before. In response to some criticisms by Ernest Gellner ([1956], pp. 252–3),[1] John Watkins ([1957], p. 170) allows that statistical regularities in social life are exceptional in not lending themselves to individualistic explanation. Here is a case 'where methodological individualism does not work'.

Watkins's retreat from strict individualism in the face of Gellner's criticism seems surprising and unnecessary. It is backed by what seems to me an erroneous argument which proceeds as follows:

> Suppose I successively place 1000 individuals facing north in the centre of a symmetrical room with two exits, one east the other west. If about 500 leave by one exit and about 500 by the other I would not try to explain this in terms of tiny undetectable west-inclining and east-inclining differences in the individuals, for . . . such an 'explanation' would merely raise the further problem: why should these west-inclining and east-inclining differences be distributed approximately *equally* among the individuals? (*ibid.*, p. 170).

Watkins seems to hint at one reason why the explanation he mentions might not do and that is that the tiny and undetectable inclinations used in the explanation may have been especially dreamt up for this explanatory purpose in the absence of any further evidence for their existence. The explanation would then be *ad hoc* and rather incredible. But I don't think that this can be the main thrust of Watkins's argument since there would seem to be no reason why the factors inclining individuals east or west should be 'tiny and undetectable' or why they should not be evidenced independently of the present west- and east-inclining activities.

---

[1] Page reference is to the reprint in O'Neill [1973].

The main claim in Watkins's case for departing from methodological individualism clearly lies in the 'further problem' which any explanation for a regularity is said to provoke. However, I do not see that any extra problem does arise.

Consider an individualistic explanation for why Mr Smith emigrated to Australia last year. No doubt this will involve various beliefs about the Australian climate and economy, certain preferences and so on. Now take the case of Mr Jones who also departed for Australia last year and whose behaviour could be rationalised in individualistic terms, too. The individualist may now explain why both Mr Smith and Mr Jones emigrated last year. But then we have also explained why *at least two* people emigrated last year; and if around 500 people have made their new homes in the Antipodes in each of the last fifty years and if each of their actions was individualistically explicable then the *regularity* of about 500 emigrants *per annum* to Australia is automatically explained too. In other words, observed regularities in behaviour do not by themselves present special difficulties for methodological individualism.

It is only where no individualistic explanation is available for each member of a sequence of actions that regularities within the sequence require another sort of explanation. Similarly, it is only when the individual outcomes of the throw of a die, say, cannot be predicted from the characteristics of the throw that holistic propensities are introduced by physical scientists to explain regularities in sequences of throws. Thus, the mere appearance of some social regularity does not militate against individualism; rather it is the breakdown of individualism which argues for an irreducible propensity determining the regularity.

After having admitted that individualism does not work in the case of social regularities Watkins appears to try not to have his cake and not to eat it at the same time, for the displacement of individualism does not seem to open the way for full-blooded holism. This is what he says:

> The existence of these actuarial regularities does not, as has often been alleged, support the historicist idea that defenceless individuals like you and me are at the chance mercy of the inhuman and uncontrollable tendencies of our society. It does not support a secularised version of the Calvanist idea of an Almighty Providence who picks people at random to fill his fixed damnation quota. For we can control these statistical regularities in so far as we can alter the conditions on which they depend. For example, we could obviously abolish road accidents if we were prepared to abolish motor-traffic (*ibid.*, pp. 170–1).

Thus according to Watkins, the occurrence of irreducible statistical regularities is not such a serious departure from individualism as one might think since 'the central assumption of the individualist position . . . is that no social tendency exists which could not be altered *if* the individuals concerned both wanted to alter it and possessed the appropriate information' (*ibid.*, pp. 168–9).

However, the question of whether we could have control over a social tendency or not is a red herring in this discussion; it is irrelevant to the question of whether such tendencies actually exist as real social forces. Neither does it bear on whether the life of every individual from Adam to Armageddon is, as a matter of fact, under their sway; after all, no human being may, as a matter of fact, ever possess the information requisite to the alteration of the social tendencies. Watkins thus stops short of allowing that a statistical social regularity might arise from a ***physically real tendency*** of the whole social group, as Durkheim and other holists have maintained.

Now, one undoubted property of the whole group is its numerical size and May Brodbeck has suggested that if the behaviour of some agents were influenced by the magnitude of the society to which he belongs then such a case would instance a breakdown in individualism. To this Watkins has replied that

> there are two *kinds* of reason why Jones's behaviour might change in groups of a certain size. His response to the situation might be an understandable personal response, in keeping with his character. (Perhaps he becomes prone to claustrophobia when the hall gets too crowded.) This would be in accord with methodological individualism. But there is another kind of reason for Jones's behaviour, strange but not inconceivable. He *might* come under a new and weird sort of control when a group to which he belongs reaches a certain critical size. When it reaches this size, he and other members of the group may cease acting under their own steam and become like iron filings in something analogous to a magnetic field which organised him and them into a holistic pattern of behaviour (Watkins [1955], pp. 180–1).

While Watkins agrees that the second kind of phenomenon is conceivable he denies that it is 'believable'.

I shall argue, in contrast, that the types of social propensity which I have described may be holistic in the sense that their strengths depends on the size of the group in which they appear. The following example may perhaps make the existence of such holistic tendencies more believable. The example is taken from the standard treatment of the inbreeding phenomenon in population genetics.[1]

We consider the simplest case of a population mating in a random fashion, where no selection or mutation occurs and where the population size remains the same in each generation. We will consider a population of size N and a single locus carrying genes A and a, whose overall frequencies in the population are p and q, respectively. Under such circumstances reproduction is similar to random sampling from a Bernoulli distribution. Individuals may be thought of as contributing their genes to an urn from which a random selection of $2N$ genes is withdrawn in the production of the new generation. In general, the probability of selecting a sample containing r A-type genes and $2N$-r a-type genes is given by

---

[1] See, for example, Falconer [1960], ch. 3.

$$^{2N}C_r\, p^r\, q^{2N-r}.$$

We thus obtain a sampling distribution of the gene frequencies appearing in the filial generation. The most probable outcome of the reproduction is a filial population with the same genetic constitution as its parent population. The above formula gives the probabilities of the various possible deviations from the parent population. It can easily be shown that the standard deviation of the sampling distribution is given by

$$\left(\frac{pq}{2N}\right)^{\frac{1}{2}}$$

It follows that the tendency of the society to produce a successor society of any particular sort ***depends on the numerical size of that soceity.***

5 CONCLUSION

Durkheim's contention, on the basis of his suicide data, that societies possess holistic social forces producing, or tending to produce, fixed characteristic rates of self-destruction has often been held as unscientific and unreasonable. I have argued that this is not so; that the social forces Durkheim had in mind are a species of the irreducible propensities or objective physical probabilities which Popper has shown are, in fact, scientifically acceptable. I also show that we have good evidence in the genetic theory that social propensities do exist and vary between human groups.

These considerations refute the view that an holistic social science is impossible and suggest the investigation of holistic social propensities and their laws as a fecund area of research.

London School of Economics

REFERENCES

DURKHEIM, E. [1897]: *Suicide*, translated into English by J. A. Spaulding and G. Simpson (1952).

FALCONER, D. C. [1960]: *Introduction to Quantitative Genetics.*

GELLNER, E. A. [1956]: 'Explanations in History', *The Proceedings of the Aristotelian Society*, Supplementary Volume 30.

O'NEILL, J. (*ed.*) [1973]: *Modes of Individualism and Collectivism.*

POPPER, K. [1957]: 'The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory', in S. Körner (*ed.*): *Observation and Interpretation.*

POPPER, K. [1959]: 'The Propensity Interpretation of Probability', *British Journal for the Philosophy of Science,* **10,** pp. 25–42.

SIMON, H. A. [1977]: *Models of Discovery, Synthese Library,* vol. 114.

WATKINS, J. W. N. [1952]: 'Ideal Types and Historical Explanation', *British Journal for the Philosophy of Science,* **3.**

WATKINS, J. W. N. [1955]: 'Methodological Individualism: A Reply', *Philosophy of Science,* **22,** pp. 58–62.

WATKINS, J. W. N. [1957]: 'Historical Explanation in the Social Sciences', *British Journal for the Philosophy of Science,* **8.**

Brit. J. Phil. Sci. 31 (1980), 329–358 *Printed in Great Britain*

A New *Begriffsschrift* (II)

by JOHN P. MAYBERRY

4 *Exposition of the new* Begriffsschrift
  (*i*) *The Basic Principles of the Theory*
  (*ii*) *The Ontological Status of Functions and Concepts*
5 *Conclusions*

4 EXPOSITION OF THE NEW BEGRIFFSSCHRIFT

By the expression 'new *Begriffsschrift*' I don't mean just a formal 'language' or conceptual notation, but a system consisting of five things: an ontology, a conceptual notation, a semantical theory connecting the two, a theory of definition, and a theory of proof. Of these, the ontology is, no doubt, the most important. But without the other components of the system, the exposition of the ontology remains unsatisfactorily vague, and we should never succeed in convincing ourselves that the system could accomplish what it is intended to do, namely, to provide a logical and metaphysical foundation for mathematics.

But what should we expect of such a theory? What does 'providing a foundation' mean in this case? There is a surprising amount of confusion on this matter. One often hears it said that 'ordinary' mathematics can be 'interpreted in' set theory or 'reduced to' it. Sometimes this takes the form of the claim that the propositions of mathematics can be translated into the formal predicate language which is based on the predicates 'ϵ' and '$=$'. Such views are quite inadequate. They evoke the picture of an artificial world set up in parallel to the 'natural' world of mathematical objects in such a way that all natural mathematical notions, or, at any rate, the essential ones, can be made to correspond, in a more or less *ad hoc* manner, to suitably chosen set theoretical notions. In this way, the theorems of mathematics carry over into provable assertions about sets. But this picture is seriously misleading. There can be no question of 'translating' ordinary mathematics into set theory: each mathematical object or structure is already set theoretical in its very essence. A group, for example, simply *is* an ordered pair (G, o) where G is a non-empty set and o is a binary operation

Received 23 *July* 1979

on G satisfying the familiar conditions.[1] There is no 'natural' or 'naive' notion of group for which this set-theoretically defined notion of group goes proxy. Groups are defined in terms of binary operations, which are defined in terms of relations, which are defined in terms of Cartesian products, and so on, back through an unbroken chain of definitions ending in the primitive notions of set theory, which cannot be further defined, but must simply be grasped.

Now such definitions are not merely abbreviations, although for certain logical purposes they may be so construed. They are, in fact, the very life's blood of modern mathematics, the source of its immense superiority over traditional mathematics—say, the mathematics of the nineteenth century—both in point of rigor and, more importantly, in point of generality. Moreover, set theory is, among other things, the theory of such definitions.

It is essential that we distinguish the obviously true claim that modern mathematics is based on set theory, from the obviously absurd one that it simply is set theory. Set theory tells us how to carry out mathematical definitions; it does not tell us which definitions we should carry out. If I know only the principles of set theory I know virtually no mathematics at all. For I know none of the essential definitions which organize the subject into intelligible and mathematically significant theories. The sense in which all of mathematics is contained in the basic laws of set theory is roughly the same as the sense in which all of English literature is contained in Oxford English Dictionary.

The new *Begriffsschrift*, then, will provide a foundation for mathematics in the following sense: it will identify and classify the entities out of which the objects and structures of mathematics can be constructed, the bricks and mortar of the subject, so to speak; it will specify the grammar and syntax of the sentences which constitute mathematical discourse; it will provide a theory of meaning for such sentences; it will prescribe how new notions are to be introduced by means of mathematically rigorous definitions; and, finally, it will lay down rules for conducting proofs.

The new *Begriffsschrift* thus delimits the *forms* in which mathematics is to be expounded. But its conceptual notation is not to be regarded as confined to a fixed set of primitive symbols. On the contrary, it should be seen as a syntactical framework whose vocabulary is indefinitely extendable by means of precisely controlled definitions. Its vocabulary grows as mathematics itself grows. The conceptual notation can even accommodate the introduction of new undefineable primitive notions should the need arise, for example, in the application of mathematics to other scientific disciplines.

---

[1] Of course one sometimes says that a group is a set *equipped* with a binary operation such that. . . . But how do you 'equip' a set with a binary operation? How can you tell when a set is so 'equipped'? *etc.*

The new *Begriffsschrift* is not a rival to 'ordinary' mathematics, but a precise formulation of it. Unlike the original *Begriffsschrift* it is essentially conservative in intention, at least in its Cantorian version. And this reflects the fact that present day mathematicians have, as Frege's contemporaries had not, a clear and coherent conception of what mathematics is about. The conservatism of the new *Begriffsschrift* is manifested in the fact that the set-theoretical principles incorporated in it are those of the Zermelo–Fraenkel system. For the Zermelo–Fraenkel notion of set is the notion employed 'naively' by contemporary mathematicians. And this is simply because it is the only plausible, or, indeed, intelligible, notion of set which anyone has ever suggested. It has no serious rivals.[1]

4(*i*) THE BASIC PRINCIPLES OF THE THEORY

The ontology of the new *Begriffsschrift* rests upon the fundamental Fregean category of *objects*. This category is characterised by the fact that it encompasses everything to which the classical relation of identity can be applied: if 'a' and 'b' denote objects, then the sentence '$a = b$' is determinately either true or false. Objects are thus sharply individuated: each object has sharp boundaries, as it were, and is essentially distinguished from every other object. Objects may therefore be said to be finite, in the primary sense of 'finite' explained in section 3(*ii*).

The category of objects is subdivided into *sets* and *individuals*. The notion of set I have already discussed at length. A set is a quantitatively determined plurality, a definite number of (definite) objects. The inclusion of sets in the category of objects can be justified because the identity criterion which applies to them is especially simple: sets are identical when they have exactly the same elements.

An individual is any object which is not a set. And that is all, strictly speaking, that needs to be said on the matter. No further specification of what is to count as an individual is required. Nor would any be desirable. For we are concerned with the foundations of mathematics. And this demands that our theory should be ontologically neutral, insofar as that is possible. We require only that individuals, whatever they may be, are properly individuated objects satisfying the classical laws of identity. Moreover, we must insist that all genuine entities which satisfy these laws, and are not sets, should count as individuals, however much they may differ from other individuals in other respects.

---

[1] On this point see Kreisel [1969], [1967], and the appendix to Kreisel and Krivine [1967], or Gödel [1947]. The discussion in Part I of my [1977] is also relevant here.

Of course we may be tempted to ask what it is that makes an individual the sort of object it is (*e.g.* a physical body, if such there be), or to search for a metaphysical *principium individuationis* which can account for the fact, taken here simply as a datum, that each individual is sharply distinguished from every other object. Who can say where such questions might lead? Perhaps some may wish to employ the scholastic apparatus of quiddities and haecceities to explain these things. But we are not called upon to provide explanations of this sort here. Such metaphysical entities, if, indeed, there are such things, must humbly take their place along side the other objects of our theory.[1]

As regards objects, then, our ontology is a minimal one, minimal, at least, consistent with accounting for the employment of sets in mathematics. And the minimality of our ontology makes it metaphysically neutral. It can accommodate the most extreme ontological profligacy or the most astringent nominalism.

Our ontology also includes the category of functions.[2] Functions are either of *first level*, taking objects as arguments, or of *second level*, taking both[3] objects and functions of first level as arguments. As in Frege's original theory a function is always universally defined: thus a function defined for one argument of a given type (*e.g.* an object) is defined for all arguments of that type. For functions of first level this has the consequence that their domains of definition are not classical domains. As I explained in section 3(*iii*), this entails that such functions cannot satisfy classical identity conditions. They are not, therefore, properly individuated, and consequently must be strongly distinguished from objects. Here, rather than in their 'unsaturatedness' lies the explanation for the necessity of excluding them from the category of objects.

The primitive signs, complete and incomplete, of the new *Begriffsschrift* are given in Figure 1. In order to emphasise the continuity between our new system and Frege's original one, I have listed the primitive signs of the new system in parallel with those of the old, and have retained Frege's notation so as to facilitate comparison. Where a sign is common to both systems, the stipulation of its reference is the same, with the exception of the sign '$\backslash\xi$', where the relation is only that of analogy.

The references of the signs '$\xi \in \zeta$', '$\xi \subseteq \zeta$', '$P(\xi)$', '$\bigcup(\xi)$', '$\{x \in \xi : \phi(x)\}$', and '$\{f(x) : x \in \xi\}$' have been laid down so that their values at arguments which are either individuals or sets are fully determined. Similarly the references of the second level function signs '$\{x \in \xi : \phi(x)\}$' and '$\{f(x) : x \in \xi\}$' are fully determined at all first level function arguments.

---

[1] Unless, of course, we take them to be functions or concepts.

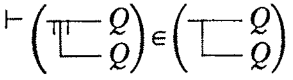
[2] Among these we include, for the time being, concepts (but see the discussion below).

[3] Frege would have described these as being of *mixed* level.

The new *Begriffsschrift*, expounded in this way has the same simple and direct referential semantics that Frege's original system had. But it also shares one important drawback, namely the difficulty, discussed in section 2(*iii*), concerning those two peculiar objects, the True and the False. We simply do not know what these objects are. Their essential natures escape us.

Now there seem to be three ways to face the problem of truth values. One is to maintain, brazenly, and in the face of all common sense, that *of course* truth values are objects, that they are what they are, and that is an end of the matter. This would appear to be the course which Frege adopts in the article 'On Sense and Reference'. But it should not recommend itself to us here. For it smacks of that very 'inventive caprice' which Frege quite rightly condemns elsewhere in his writings.

A second possible solution is to adopt a convention whereby the truth values are simply identified with previously given, well-defined objects. This is what Frege attempts to do in the *Grundgesetze*. But, as I pointed out in section 2.3, the objects which he chooses to identify with the truth values are every bit as mysterious as the truth values themselves, every bit as opaque in their essential natures. We need not follow Frege's example in this, however. We may, for example, identify the True with the Von Neumann natural number $1(=\{\varnothing\})$ and the False with $0(=\varnothing)$—no problem about the nature of the *definiens* in either case here. But there is a certain undesirable artificiality. It will, for example, turn out to be the case that



(in more conventional notation: $\vdash [\neg Q \wedge Q] \in [Q \supset Q]$). Indeed, if we adopt this convention we can dispense with the primitive object name 'ϕ' in the finitary version of the theory (though not with 'ω' in Cantorian version), since any false sentence will name the empty set.

	Frege's system	The new *Begriffsschrift*
Object names		1. $\varnothing$ (the empty set) 2. ω (the set of von Neumann natural numbers): the members of ω are $\varnothing$, $\{\varnothing\}$, $\{\varnothing, \{\varnothing\}\}, \ldots,$ $n, n \cup \{n\}, \ldots$

First level function names	1. $\xi = \zeta$ (identity): $a = b$ is (the) True if a is identical with b, and is (the) False otherwise. 2. $-\xi$ (the horizontal): $-a$ is the True if a is the True, and is the False otherwise. 3. $\top\xi$ (negation): $-a$ is the False if a is the true, and is (the) True otherwise. 4. $\begin{array}{l}\top\zeta\\ \llcorner\xi\end{array}$ (implication): $\begin{array}{l}\top b\\ \llcorner a\end{array}$ is the False if a is (the) True and b is not (the) True, and is (the) True otherwise. 5. $\backslash\xi$: $\backslash a$ is b if there is a first level concept $\phi(\xi)$ under which b, and b alone, falls, and a is $\acute{\varepsilon}\phi(\varepsilon)$; otherwise $\backslash a$ is a itself.	1. $\zeta = \xi$ (identity) 2. $-\xi$ (the horizontal) 3. $\top\xi$ (negation) 4. $\begin{array}{l}\top\zeta\\ \llcorner\xi\end{array}$ (implication) 5. $\backslash\xi$: $\backslash a$ is b if a is the set whose only member is b; $\backslash a$ is a itself otherwise. 6. $\xi \in \zeta$ (membership): $a \in b$ is (the) True if b is a set and a is a member of b; otherwise it is (the) False. 7. $\xi \subseteq \zeta$ (inclusion): $a \subseteq b$ is (the) True if a is the empty set or an individual, or if a and b are sets and every member of a is a member of b. 8. Set (ξ) (sethood): Set (a) is (the) True if a is a set, and is (the) False otherwise. 9. $P(\xi)$ (power set): $P(a)$ is the set of all sets which are included in a. 10. $\{\xi, \zeta\}$ (pair set): $\{a, b\}$ is the set whose members are a and b. 11. $\bigcup(\xi)$ (union): $\bigcup(a)$ is the set of all those objects which are members of elements of a.
Second level function names	1. $\acute{\varepsilon}f(\varepsilon)$: $\acute{\varepsilon}\mathrm{f}(\varepsilon)$ is the value-range of the function $\mathrm{f}(\xi)$. 2. $-\!\overset{\mathfrak{a}}{\smile}\!-\ \phi(\mathfrak{a})$ (the universal quantifier): $-\!\overset{\mathfrak{a}}{\smile}\!-\ \varphi(\mathfrak{a})$ is (the) True if the function $\varphi(\xi)$ is (the) True for every argument; otherwise it is (the) False.	1. $\{x \in \xi : \phi(x)\}$ (comprehension): $\{x \in a : \varphi(x)\}$ is the set of all elements of a at which the function $\varphi(\xi)$ is (the) True. 2. $\{f(x) : x \in \xi\}$ (replacement): $\{\mathrm{f}(x) : x \in a\}$ is the set of all values of the function $\mathrm{f}(\xi)$ at arguments which are elements of a. 3. $\mathrm{Rec}_x(f(x), \xi, \zeta)$ (recursion): this function will be described later.

Third level function names	1. $-\!\overset{\mathfrak{b}}{\smile}\!-M_\beta(\mathfrak{b}(\beta))$ (the second order universal quantifier): $-\!\overset{\mathfrak{b}}{\smile}\!-\mathrm{M}_\beta(\mathfrak{b}(\beta))$ is (the) True if $\mathrm{M}_\beta(\mathrm{f}(\beta))$ is (the) True for every first level argument $\mathrm{f}(\xi)$; otherwise it is (the) False.

Figure 1

Such oddities also arise in the original *Begriffsschrift*, of course, and are unexceptionable from a narrowly legalistic point of view: after all, everything is being done in accordance with the letter of our semantical stipulations, which, however strange, are still perfectly precise. But such legalism is somehow unsatisfactory here. We are dealing with matters which are too fundamental and too serious. And our referential stipulations seem to depend on an unacknowledged, but clearly more fundamental notion of the truth and falsity of the sentences of our theory.

We could avoid some of the awkwardness here if we dropped the identification of sentences and terms at the syntactical level. This would require that we make a fundamental distinction in the logical grammar between those complete signs which denote truth values and other complete signs. This would entail dropping the Fregean 'horizontal' function, $-\xi$, for a start, and then restricting the arguments in implication and negation to truth value names (sentences) or truth value marks[1] (sentence marks). It would also be necessary to distinguish concept signs from other incomplete signs in order that the comprehension and replacement functions are applied only to functions of the appropriate sort. All these modifications would be confined entirely to the logical grammar and syntax of the theory: no adjustments need be made in the ontology, or semantics, or even in the convention identifying the truth values with 0 and 1. In this way we could ignore the odd consequences of that convention simply because the propositions expressing those consequences couldn't be expressed in the modified conceptual notation.

One must acknowledge, however, that all these complications stem from the fact that truth and falsity are not, after all, objects, but properties of sentences. And this suggests the third, and from the orthodox Fregean standpoint, most radical, solution to our difficulty. This consists in simply abandoning the view that sentences are names of truth values. From this new standpoint, sentences are a fundamentally different sort of complete sign from ordinary object names, or *terms*, as they could now be called. A sentence doesn't *name*, but *has*, a truth value. We should have to acknowledge a new kind of incomplete sign, the *concept sign*, in addition to ordinary function signs. And this would entail acknowledging a new ontological category of *concepts*, which, like functions, would occur at different levels depending on the types of their arguments. In accordance with this approach, for example, the comprehension function, $\{x \in \xi: \phi(x)\}$, is a second level function taking a unary first level *concept* and an object as arguments.

---

[1] Frege calls a complete term containing free variables (or 'Roman letters' in his terminology) an 'object mark'.

The resulting ontology is somewhat more complicated than the strictly orthodox one in that it recognises more types of insaturata. But it still permits us to enjoy most of the benefits of the Fregean referential semantics, without the inconvenience of positing truth values as objects. Let us turn, then, to a detailed exposition of the new *Begriffsschrift* based on this third approach.

The fundamental ontological categories of the theory are the category of *objects*, the category of *functions*, and the category of *concepts*. Objects I have already discussed at some length. The "unsaturated" entities, functions and concepts, both differ from objects in that they lack clear-cut identity criteria. The distinction between functions and concepts consists in this: a function *applies to* its arguments to *yield* a value; a concept *is true of* or *is false of* its arguments. This distinction is mirrored in the grammar of the conceptual notation. For example, the name of a first level function and the name of an object combine syntactically to produce a name of the object which is the value of the given first level function at the given object as argument. Similarly, the name of a first level concept and the name of an object combine syntactically to produce a sentence which is true, or false, according as the given concept is true of, or false of, the given object.

The primitive signs of the modified conceptual notation are those listed in Figure 1, with the exception of the sign '$-\xi$' designating Frege's 'horizontal' function. We must, however, make a fundamental distinction between function names and concept names. Moreover the signs '$\begin{array}{l}\top\!\!-\!\!-\zeta\\ \llcorner\!\!-\!\!-\xi\end{array}$' and '$-\!\!\top\!\!-\xi$' are longer to be regarded as names at all, and their argument places must always be filled by sentences. In fact, I shall employ the more conventional notations '$\supset$' and '$\neg$' in their stead, and shall also use the familiar signs '$\wedge$' (and), '$\vee$' (or), and '$\equiv$' (if, and only if), for the other commonly employed propositional connectives.

The letter 'x' occurring in the signs '$\{x \in \xi: \phi(x)\}$' and '$\{f(x): x \in \xi\}$' is a *binding letter*. In general, it will be necessary to employ several distinct binding letters in an expression to avoid the familiar problem of unintended binding. Thus in any given context it must be made clear, implicitly or explicitly, what symbols are being used as binding letters. I shall usually employ the letters 'x', 'y', and 'z', with or without subscripts, for this purpose.[1]

Finally, our primitive vocabulary must include the humble brackets:

'[' and ']'

There are two kinds of complete expressions employed in the conceptual notation: *terms* and *sentences*. Terms *denote* objects; sentences *have* truth values, true or false. Terms and sentences are formed in the obvious manner from the primitive symbols of the system.

By virtue of the semantical stipulations laid down in Figure 1, all of the terms of our conceptual notation have references and all of the sentences have truth values. But we have no mechanism by which to express generality. For this we need free variables or *parameters*. Three sorts of parameters may be required: *object parameters*, *function parameters* and *concept* or *predicate parameters*. In any given context in which parameters are employed, it must be clearly indicated, implictly or explicitly, which symbols are being employed as parameters, what types of parameter they are, and, in the case of function and predicate parameters, what their degrees are, *i.e.*, how many arguments they require. Naturally, the symbols employed as parameters in a given context must be distinct from each other, from the primitive symbols of the conceptual notation, and from the binding letters used in that context. I shall generally use the letters 'a', 'b', and 'c', with or without subscripts as object parameters, the letters 'F', 'G', and 'H', with or without subscripts, to form function parameters '$F(\xi)$', '$G(\xi, \zeta)$', *etc.*, and the letters 'Q', 'R' and 'S', with or without subscripts, to form predicate parameters '$Q(\xi)$', '$R(\xi, \zeta)$', *etc.* When parameters are present we speak of *parametric terms*, *parametric sentences*, and, eventually, of *parametric proofs*. These must be distinguished from terms proper, sentences proper, and proofs proper.

In the *Grundgesetze* Frege employs the judgment stroke, '$\vdash$', to convert a mere name of a truth value into a judgment or proposition. Since in the present case sentences are not mere names, we have no need for such a device. In modern treatments of logic this sign is often used to indicate formal provability: thus '$\vdash A$' stands for 'There is a formal proof whose last line is A'. I shall not employ the sign in this way either, however, for I have yet another use for it. I shall take '$\vdash A$' to mean ' "A" is formally valid', and I shall take *formal validity* to be a primitive logical notion which cannot be defined, although, of course, it can be informally explicated.

If 'A' is a sentence proper, then to say that 'A' is formally valid is simply to say that it is true. If '*A*' is a parametric sentence, then to say that '*A*' is formally valid entails that any sentence proper, in any legitimate extension of the basic conceptual notation, which has the form of '*A*', that is to say, which could be obtained by substituting non-parametric expressions of the appropriate sort (complete or incomplete) for all the parameters of '*A*', would be true.[1]

---

[1] I shall also regard expressions in the conceptual notation as being the same, if they are *alphabetic variants* of one another, *i.e.* if they differ only in their binding letters.

An expression of the form '$\vdash A$', employed assertively, I shall call a *formal proposition* of the new *Begriffsschrift*. This is by way of contrast with an *ordinary proposition* which is a sentence proper, '*A*', employed assertively. A parametric sentence cannot, of course, ever be taken to be a proposition at all, since its parameters have no determinate references. Formal propositions are not propositions in the sense of classical logic, for they are neither true nor false, but only *justified* or *unjustified*. A formal proposition cannot be a classical proposition because it has no negation, at least not in the strictly classical world of the new *Begriffsschrift*. For to say that a formal proposition is unjustified is not to make a mathematically significant statement, but merely to comment on the state of mathematical knowledge.

Just as we must distinguish between '*A*' and ' "*A*" is true' (which can be rendered '$\vdash A$'), so we must also distinguish between '$\vdash A$' and ' "$\vdash A$" is justified'. In the new *Begriffsschrift* sentences proper can be *used*, in ordinary propositions, or *mentioned* in formal ones. Expression of the form '$\vdash A$', however, can only be used, never mentioned, at least not in contexts such as ' "$\vdash A$" is justified' which are the logically significant ones. Moreover, a formal proposition can never occur as a constituent of a complex proposition, but must always be used assertively, and on its own.

All of these apparently otiose conventions and distinctions are, unfortunately, really necessary. They are forced upon us by the logical and semantical difficulties surrounding global or general assertions in set theory. These, in turn, arise from the metaphysical problem of the infinite discussed in section 3(*ii*). They are unavoidable. To tackle them directly, we should have to follow the path cut by Brouwer. But that would take us beyond the boundaries of classical logic and mathematics, and in this essay, I am conducting my expedition only as far as the frontiers of that familiar and extensively cultivated territory.

Before I proceed to an account of definition and proof, I want to reiterate what I said about the nature and purpose of the conceptual notation, namely, that it is not to be regarded as a closed or complete system with a fixed stock of symbols, but rather as a general syntactical framework whose vocabulary is indefinitely extendable. This is why I have treated binding letters and parameters as auxiliary symbols outside the conceptual notation proper, brought into play, in whatever symbolic guise may prove convenient, as the occasion demands.[1] I shall take a similar view of the conventions governing definition and proof. In the case of definitions, we can lay down, once and for all, complete and adequate stipulations governing their employment. But this is notoriously not possible for proofs. There our stipulations can only be partial, and consequently allowance must be made for extending them. Moreover, it is desirable that considerable flexibility be permitted for the concrete realisations of our conventions and stipulations. I shall hold it out as an ideal, though perhaps unrealizable, aim that it should be possible so to enlarge our vocabulary and to augment our conventions of proof, that formal arguments in the conceptual notation might approach the perspicuity and simplicity of informal arguments in everyday mathematical language, but without suffering from any of the vagueness and imprecision which can accompany such informality.

---

[1] My use of the word 'entails' here, and of the subjunctive mood of the verb ('*would* be true') betray the fact that I am not giving even an informal definition of formal validity, but am merely indicating the normative consequences, as it were, of an assertion of formal validity. For the purposes of the new *Begriffsschrift* I shall regard such assertions as being *sui generis*, and, ultimately, unsusceptible to paraphrase.

The overt claims of our theory take the form of ordinary propositions which are claimed to be true, or formal propositions which are claimed to be justified. The *basic laws*, or *axioms*, of the theory are all formal propositions. They fall naturally into seven fundamental groups.

I. Basic Laws of Identity

1. $\vdash a = a$ (*Reflexive Law of Identity*)
2. $\vdash a = b \supset b = a$ (*Symmetric Law of Identity*)
3. $\vdash a = b \supset [Q(a) \equiv Q(b)]$ (*Law of Substitutivity of Identity*)

II. Basic Laws of Elementary Set Theory

1. $\vdash \mathrm{Set}\,(a) \equiv [a \subseteq \varnothing \supset a = \varnothing]$
2. $\vdash a \subseteq a$ (*Reflexive Law of Inclusion*)
3. $\vdash [a \subseteq b \wedge b \subseteq c] \supset a \subseteq c$ (*Transitive Law of Inclusion*)
4. $\vdash [a \subseteq b \wedge c \in a] \supset c \in b$
5. $\vdash [\mathrm{Set}\,(a) \wedge \mathrm{Set}\,(b)] \supset [[a \subseteq b \wedge b \subseteq a] \supset a = b]$ (*Law of Extensionality*)
6. $\vdash \varnothing \subseteq a$

III. Basic Laws Characterizing the Fundamental Functions

1. $\vdash a \in P(b) \equiv [\mathrm{Set}\,(a) \wedge a \subseteq b]$
2. $\vdash a \in \{b, c\} \equiv [a = b \vee a = c]$
3. $\vdash a \in \{x \in b\colon Q(x)\} \equiv [a \in b \wedge Q(x)]$
4. $\vdash \mathrm{Set}\,(\{x \in a\colon Q(x)\})$

---

[1] This convention also makes it possible actually to use the conceptual notation, whereas more conventional formulations of formal syntax require that the system be expounded in the meta-theory using informal syntactic variables.

In order to state the remaining basic laws of group III in a convenient and perspicuous form, we must first lay down some definitions.

Definitions of the Universal and Existential Quantifiers

1. $\vdash (\forall x \in a)\,[Q(x)] \equiv a \subseteq \{x \in a\colon Q(x)\}$
2. $\vdash (\exists x \in a)\,[Q(x)] \equiv \neg\,(\forall x \in a)\,[\neg Q(x)]$

The quantifiers $(\forall x \in \xi)\,[\phi(x)]$ and $(\exists x \in \xi)\,[\phi(x)]$ are second level concepts. The definition of a concept always takes the form of a formal proposition in which the sign '$\vdash$' is attached to a parametric sentence whose principal logical sign is '$\equiv$'. The right hand side of the equivalence is the ***definiens***, the left hand side the ***definidendum***. The ***definiendum*** consists of the defined concept sign applied to distinct parameters of the appropriate types as arguments (binding letters counting as parts of the concept sign), and the ***definiens*** consists of a parametric sentence all of whose non-parametric signs are either primitive or previously defined, and all of whose parameters occur in the ***definiendum***.

III. (continued)

5. $\vdash a \in \bigcup(b) \equiv (\exists x \in b)\,[a \in x]$
6. $\vdash \mathrm{Set}\,(\bigcup(a))$
7. $\vdash a \in \{F(x)\colon x \in b\} \equiv (\exists x \in b)\,[a = F(x)]$
8. $\vdash \mathrm{Set}\,(\{F(x)\colon x \in a\})$
9. $\vdash a = \backslash b \equiv [[[\neg \mathrm{Set}\,(b) \vee (\exists x \in b)\,[(\exists y \in b)\,[\neg[x = y]]]] \wedge a = b] \vee [a \in b \wedge (\forall y \in b)\,[a = y]]]$

For the Cantorian version of our theory we need a basic law characterising the set ω of von Neumann natural numbers. But for this we must first define the concepts Ord(ξ) (ξ is a von Neumann ordinal) and Finite (ξ) (ξ is a finite set).

Definitions of the Concepts of Ordinal Number and Finite Set

1. $\vdash \mathrm{Ord}\,(a) \equiv \mathrm{Set}\,(a) \wedge (\forall x \in a)\,[x \subseteq a] \wedge (\forall x \in a)$ $[(\forall y \in a)\,[\neg[x = y] \supset [x \in y \vee y \in x]]]$
2. $\vdash \mathrm{Finite}\,(a) \equiv \mathrm{Set}\,(a) \wedge (\forall x \in P(P(a)))\,[\neg[x = \varnothing] \supset$ $(\exists y \in x)\,[(\forall z \in x)\,[z \subseteq y \supset z = y]]]$

The definition of the von Neumann ordinals is that of R. M. Robinson; the definition of finite set is Tarski's. There is no need to include a wellfoundness condition in the definition of ordinal number since the wellfoundness of the membership relation can be derived from the Axiom of Foundation given below.

IV. The Axiom of Infinity

$$\vdash a \in \omega \equiv [\mathrm{Ord}\,(a) \wedge \mathrm{Finite}\,(a)]$$

Definition of the Intersection Function

$$\vdash a \cap b = \{x \in a \colon x \in b\}$$

This illustrates the general method for defining functions. Such a definition always takes the form of a formal proposition in which the sign '$\vdash$' is attached to an identity. The right hand term of the identity is the *definiens*, the left hand term the *definiendum*. The latter consists of the function sign to be defined applied to the appropriate number of distinct parameters of the required types as arguments. The *definiens* is a term all of whose non-parametric signs are either primitive or previously defined, and all of whose parameters occur in the *definiendum*.

V. The Axiom of Foundation

$$\vdash [\mathrm{Set}\,(a) \wedge \neg[a = \varnothing]] \supset (\exists x \in a)\,[x \cap a = \varnothing]$$

The Axiom of Foundation says that the membership relation is well-founded. It reflects the fact that there can be no infinite descending membership chains

$$\ldots \in \alpha_{n+1} \in \alpha_n \in \ldots \in \alpha_1 \in \alpha_0$$

If such a chain were to exist, we could form no conception of its constituents as sets. For the essence of the set whose members are $\alpha, \beta, \ldots$ is that its members *are* $\alpha, \beta, \ldots$, and that these are all well-defined objects. The existence of an infinite descending membership chain would thus violate the Aristotelian principle that there can be no infinite regress in formal causes or definitions.[1]

The Axiom of Foundation provides the basis for the familiar global description of the set-theoretical universe as a *cumulative hierarchy* in which all objects occur at *levels* indexed by ordinal numbers. At level 0 occur all those objects which have no members, namely, the empty set and the individuals. At level $\alpha+1$ occur all those sets whose members occur at level α. The levels are cumulative in that objects occurring at a given level also occur at all higher levels. At levels indexed by ordinals, such as ω, which have no immediate predecessor occur all those objects which occur at previous levels. To establish these facts in a mathematically rigorous way, however, we must employ the recursion function, which is described below.

---

[1] I have discussed this issue in greater detail in my article 'On the Consistency Problem for Set Theory' ([1977], part I, section 1.8, pp. 30–32).

For the Cantorian version of our theory we need the familiar Axiom of Choice.

VI. The Axiom of Choice

$$\begin{aligned}&\vdash [(\forall x \in a)\,[\mathrm{Set}\,(x) \wedge \neg[x = \varnothing]] \wedge (\forall x \in a)\,[(\forall y \in a)\,[\neg[x = y]\\ &\quad \supset x \cap y = \varnothing]]] \supset (\exists z \in P(\textstyle\bigcup(a)))\,[(\forall x \in a)\,[(\exists y \in z)\\ &\qquad [(\forall u \in z)\,[u \in x \equiv u = y]]]]\end{aligned}$$

If we assume all sets to be finite then this axiom can be derived from the other axioms as a theorem.

In order to state the remaining axiom which characterizes the recursion function, we must first lay down some further definitions.

Definitions of the Singleton, Ordered Pair and Ordered Triple Functions

1. $\vdash \{a\} = \{a, a\}$
2. $\vdash (a, b) = \{\{a\}, \{a, b\}\}$
3. $\vdash (a, b, c) = ((a, b), c)$

Definition of the Boolean Union Function

$$\vdash a \cup b = \textstyle\bigcup(\{a, b\})$$

Definition of the Concept of a (Local) Relation.

$$\vdash \mathrm{Rel}\,(a) \equiv (\forall x \in a)\,[(\exists y \in \textstyle\bigcup(\bigcup(a)))\,[(\exists z \in \bigcup(\bigcup(a)))\,[x = (y, z)]]]$$

("Rel (ξ)" should be read "ξ is a (local) relation")

Definitions of the Domain, Range and Field of a (Local) Relation

1. $\mathrm{Dom}\,(a) = \{x \in \bigcup(\bigcup(a)) \colon (\exists y \in \bigcup(\bigcup(a)))\,[(x, y) \in a]\}$
2. $\mathrm{Range}\,(a) = \{x \in \bigcup(\bigcup(a)) \colon (\exists y \in \bigcup(\bigcup(a)))\,[(y, x) \in a]\}$
3. $\mathrm{Field}\,(a) = \mathrm{Dom}\,(a) \cup \mathrm{Range}\,(a)$

Definition of the Concept of a Well-Founded (Local) Relation

$$\begin{aligned}&\vdash Wf\,(a) \equiv \mathrm{Rel}\,(a) \wedge (\forall x \in P(\mathrm{Field}\,(a)))\,[\neg[x = \varnothing]\\ &\quad \supset (\exists y \in x)\,[(\forall z \in x)\,[(z, y) \in a \supset z = y]]]\end{aligned}$$

('$Wf(\xi)$' should be read 'ξ is a well-founded (local) relation')

Definition of the Initial Segment Function

$$\begin{aligned}&\vdash \mathrm{Seg}\,(a, b) = \{x \in \mathrm{Field}\,(b) \colon [Wf(b) \wedge a \in \mathrm{Field}\,(b)]\\ &\quad \wedge [\neg[x = a] \wedge (x, a) \in b]\}\end{aligned}$$

('Seg (ξ, ζ)' should be read 'The initial segment of ζ determined by ξ' or simply 'The ξ-segment of ζ'. Notice that if b is not a well-founded relation or if a is not in the field of b, then Seg (a, b) is the empty set.)

The name of the recursion function in the conceptual notation is '$\mathrm{Rec}_x(f(x), \xi, \zeta)$'. The function is of second level, taking a unary first level function and two objects as arguments. The letter 'x' occurring in the name is a binding letter.

VII. The Axiom of Recursion

$$\vdash \mathrm{Rec}_x(F(x), a, b) = F((\{(z, \mathrm{Rec}_x(F(x), a, z)) : z \in \mathrm{Seg}\,(a, b)\}, a, b))$$

('$\mathrm{Rec}_x(f(x), \xi, \zeta)$' should be read 'Recursion with respect to x in $f(x)$ along ξ evaluated at ζ'.)

The recursion function permits us to introduce a function $\mathrm{f}(\xi)$ satisfying the recursion equation

$$\mathrm{f}(a) = \mathrm{g}((\mathrm{f}/\!\!\succ_r a, r, a))$$

where $\mathrm{g}(\xi)$ is a given function and r is a well-founded local relation. In such circumstances, $\mathrm{f}/\!\!\succ_r a$ is the graph of the restriction of the function $\mathrm{f}(\xi)$ to the predecessors of a in the relation r, in symbols:

$$\mathrm{f}/\!\!\succ_r a = \{(x, \mathrm{f}(x)) : x \in \mathrm{Seg}\,(a, r)\}$$

If r is not a well-founded relation, or if a is not in the field of r, then $\mathrm{f}(a)$ so defined is equal to $\mathrm{g}((\phi, r, a))$. I shall discuss the recursion function and its rôle in the new *Begriffsschrift* in section 4(*ii*).

This completes the exposition of the basic laws of our system. We must now consider its proof theory. The fundamental distinction is that between ***proofs proper*** and ***proof forms*** or ***parametric proofs***. I shall use the term 'proof', unqualified, for both of these notions. A proof consists of an ordered series of ***steps***. In a proof proper each step consists of the assertion of a formal proposition, or of an ordinary proposition, or of the exhibition of a parametric proof. The assertions occurring in a proof proper must either be basic laws, definitions, or the conclusions of instances of rules of inference whose premisses occur as earlier steps in the proof. If a defined symbol occurs in a given step, its definition must occur as an earlier step. A parametric proof is like a proof proper except that a parametric sentence can stand in the place of a sentence proper. A parametric proof is thus the form of a proof proper in the same way that a parametric sentence is the form of a sentence proper. The whole of a parametric proof stands, as it were, between inverted commas: even the formal propositions and sentences proper occurring in it are not used assertively. When a parametric proof occurs as a step in another proof, its parameters are not parameters of the larger proof, nor are parameters occurring in formal propositions parameters of proofs in which those propositions occur.

There are four basic rules of inference in our theory.

Rule 1. From $\vdash A$ as premiss infer A', where A' is any substitution instance of A.

Rule 2. From $A_1, \ldots, A_n$ $(n \geqslant 0)$ as premisses, infer A, where A is a truth functional consequence of $A_1, \ldots, A_n$.

Here the inclusion of the case $n = 0$ allows immediate assertion of a tautology.

Rule 3. From a (parametric) proof of A as premiss, infer $\vdash A$.

Rule 4. From $\vdash A \supset [p \in s \supset p \in t]$ as premiss, where p is a parameter not occurring in A, s, or t, infer $\vdash A \supset [s \subseteq t]$.

We may allow considerable scope for variation in the actual presentation of proofs. All that is required is that it be specified what the steps of the proof are, and what is their order of precedence. The latter need not be linear, as long as the premisses always precede the conclusion in any application of a rule of inference, and the step which represents the conclusion follows all the other steps of the proof. The conclusion must, of course, always be a proposition, formal or ordinary, in the case of a proof proper, or a parametric sentence, in the case of a parametric proof.

The basic laws and rules of inference of our theory are not, of course, complete. They can be extended indefinitely. Indeed, in the interest of efficiency it might be desirable to augment our rules, for example by allowing proofs from hypotheses. Nevertheless, the system of proofs as it stands suffices to establish all known mathematical theorems.

The one feature of the system which represents a significant departure from customary usage is the meaning given to the Fregean judgment stroke, '$\vdash$', the familiar "assertion sign". Because our conceptual notation lacks, indeed, must necessarily lack, global quantifiers, there must be a formal mechanism for formulating global propositions. To be sure, we could get by without this if we were prepared to have as our basic laws, not the formal propositions we have taken, but the ordinary propositions which are their substitution instances. But it seems to me that these laws are binding not only on the propositions expressed by sentences formed using the limited basic vocabulary of our conceptual notation, but on all possible propositions which could be formed by arbitrary extension of that notation. The problem of general or global propositions in set theory won't go away simply because we refuse to face it.

The notion of formal validity cannot be captured by any set of axioms and rules. Indeed, as I have already pointed out, it is not even a classical notion. The species of formally valid parametric sentences expressible in our basic vocabulary provides an example of a sub*species* of a set (in this case, the set of all parametric sentences expressible in that vocabulary) which does not form a set. This illustrates the general fact that a species may be indefinite without being indefinitely large.

It only remains for me to discuss the manner in which our system of proofs, and, in particular, Rule 2 determines the semantic role of the propositional connectives. This, however, is best left to the next section in which I shall turn again to the general question of the ontological status of unsaturated entities.

4(*ii*) THE ONTOLOGICAL STATUS OF FUNCTIONS AND CONCEPTS

The syntactical relation between a propositional connective and the sentences which it combines into a compound sentence is obviously similar to the syntactical relation between a function name and the object names which occupy its argument places. But if this purely grammatical analogy is to have a parallel in the ontology of the theory, it will be necessary to regard sentences as having a semantical role analogous to that of object names, that is to say, they must be taken to have references. And the only plausible candidates for the references of sentence are truth values. As Frege points out in 'Sense and Reference' the 'thought' expressed by the sentence belongs to the realm of sense not of reference. And, as I have argued, we have no cause to hypostatize senses, to regard them as objects of a peculiar kind, so they could not serve as the references of sentences in any case. But to posit truth values as objects, though less extravagant than positing senses, is to countenance as objects things of whose essential natures we can form no idea whatsoever. This is why I tentatively suggested in section 4(*i*) that if we wish to retain Frege's referential account of sentences and propositional connectives, we should adopt a convention whereby, say, the truth values are identified with the von Neumann ordinals 0 and 1. The propositional connectives could then be assigned ordinary functions as their references, and concepts could be identified with their associated characteristic functions. We could even avoid facing the odd consequences of our conventions (*e.g.* that $\varnothing \in [\varnothing = \varnothing]$) by retaining a syntactical distinction between sentences and terms so that the sentences expressing these consequences could not be formed. I rejected this approach only because it seemed to me that issue was too fundamental to settle by adopting an arbitrary convention. Such a solution seems parasitic on a more fundamental notion of truth and falsity as properties of sentences. I was forced, therefore, to distinguish concepts from functions as a different category of 'unsaturated' entities.

Now the view of propositional connectives that emerges from this alteration of the Fregean ontology and semantics is that they are syncategorematics signs, not names with references. A propositional connective simply serves to combine simple sentences into compound ones. The truth conditions of a compound sentence are reduced to those of its simple or atomic constituents by means of the truth tables for the propositional connectives occurring in it. A propositional tautology is thus necessarily true, partly by convention, namely, the convention laid down by the truth tables for the connectives, and partly by fact, namely the fact that its atomic constituents are determinately either true or false—I'm thinking here of a tautology constructed from sentences proper. However, if 'A' is a parametric sentence which is a propositional tautology, the formal proposition '$\vdash A$' is justified since only those substitutions of names for parameters are allowed which result in atomic constituents of 'A' becoming sentences with determinate truth values.

Because of the syncategorematic character of the propositional connectives, I relegated their logic to the rules of the proof theory, rather than formulating it in the basic laws. This, however, is a merely rhetorical device, affecting only the manner of presentation of the theory and not its substance.

But this treatment of the propositional connectives as syncategorematic signs naturally raises the question whether it might not be possible to treat the other incomplete signs of the system as syncategoremata as well.[1] Now it seems to me that if we restrict our consideration to the function signs treated in Axiom Group III, then such an approach in quite plausible. Each of the signs '$P(\xi)$', '$\{\xi, \zeta\}$', '$\{x \in \xi : \phi(x)\}$', '$\bigcup(\xi)$' and '$\{f(x) : x \in \xi\}$' appears in a formal proposition of the form '$\vdash A \equiv B$', where A is a parametric sentence of the form '$a \in t$' in which 't' is a parametric representation of a general term employing the function sign in question as its outermost function sign, and 'B' is a sentence whose parameters are 'a' and the parameters of 't', all of whose non-parametric signs have fixed meaning. These axioms therefore lay down necessary and sufficient conditions for membership in the references of terms formed from these function signs, and, in this sense, determine those references. In this way the axioms can be regarded, not as basic laws reflecting the way things are, but as *formal stipulations* fixing the use of incomplete signs. I say 'formal stipulations' here, rather than 'definitions' because these formal propositions do more than merely establish notational conventions. What that 'more' amounts to can perhaps best be seen by turning our attention to the Axiom of Infinity.

The Axiom of Infinity, '$\vdash a \in \omega \equiv [\mathrm{Ord}\,(a) \wedge \mathrm{Finite}\,(a)]$', has the same general form as the axioms we have just been considering. It also lays down necessary and sufficient conditions for membership in the reference of a term, in this case the set name 'ω'. Thus this axiom must also count as a 'formal stipulation'. But it is notorious that the Axiom of Infinity is not a mere definition of terms. It has genuine ontological content. For it is by means of the incorporation of the sign 'ω' in our conceptual notation, and the adoption of this axiom as a basic law, that we lay down the fundamental assumption of Cantorian set theory that the species of natural numbers is quantitatively determined, and therefore forms a set. The Axiom of Infinity embodies this assumption, and in the same way the Axiom of Power Set embodies the assumption that the species of subsets of any set is quantitatively determined. These formal stipulations thus embody the finiteness assumptions of our theory, where 'finiteness' is understood in the fundamental sense explained in section 3(*ii*).

It would thus appear possible to treat the basic function signs, not as names, but as syncategorematic signs. Just as we explain the semantical roles of the propositional connectives by giving necessary and sufficient conditions for the truth or falsity of the compound sentences constructed from them, so we explain the semantical roles of the function signs by giving necessary and sufficient conditions for membership in the references of the compound terms constructed from them.

But perhaps we should pause for a moment to consider what, on this view, are the semantical roles of the various sorts of parameters which occur in our formal propositions. Are they internal 'variables' ranging over, or 'indefinitely indicating', in Frege's turn of phrase, entities of some definite kind? Clearly we cannot regard the function and concept parameters in this way, in any event. Must we therefore regard them as 'meta-variables' whose domain is a fixed stock of symbols? And must we make a fundamental distinction between the object parameters and the higher level ones? It seems to me that we must treat all of the different sorts of parameters in the same way, but that they should not be regarded as either internal or meta-variables in the conventional way. A parameter is like a meta-variable in that it goes proxy for expressions, complete or incomplete, which have determinate 'meanings' or semantical roles. But it is unlike a meta-variable in that there is no fixed stock of expressions, laid down in

---

[1] Indeed, there is an unsatisfactory asymmetry in the version of the theory in which propositional connectives are treated syncategorematically while the other incomplete signs are taken as referring signs. This is why we should not, I believe, dismiss out of hand the orthodox Fregean version of the new *Begriffsschrift* sketched in 4(*i*).

advance, which constitute its possible substitution instances. A formal proposition is thus binding on all future extension of our basic conceptual notation.

But doesn't this mean, in effect, that we must adopt a 'substitutional' account of, say, the quantifier '$(\forall x \in P(\omega))$'? Well, I suppose, in a way, it does.[1] But it is a substitutional theory with a difference. For we do not confine the possible substitutions to a fixed vocabulary. Besides, even a strict finitist will employ quantifiers of the form '$(\forall x \in a)$' which obey the usual classical laws. '*Ah*', comes the reply, 'but if "*a*" is given a particular finite set as reference, even an extremely large one, then he can, *in theory*, enlarge his vocabulary to include names for all the members of *a*'. But I ask: in *what* theory can he do this? He cannot do it in currently accepted physical theory. He certainly cannot do it in practice. He can describe—as opposed to exhibit—an extension of the basic conceptual notation in which every member of the finite set *a* has a name. But then the Cantorian finitist can describe an extension in which every member of $P(\omega)$ has a name. It will not do for the strict finitist to protest that the set *a* is finite, for the Cantorian will merely retort that $P(\omega)$ is also finite, in the true sense of finite as opposed to the merely technical one. The argument thus comes back to a dispute over the fundamental Cantorian assumption that the species of von Neumann natural numbers forms a set. In all other respects the two opponents confront exactly analogous difficulties.

The advantage of the present approach to the ontology and semantics of the new *Begriffsschrift* over that presented in section 4(*i*) is obvious. True, the semantics is more complicated. But the ontology is much simpler. It is based on a univocal notion of being: the only mode of existence is the fundamental one attributed to objects. There is no puzzle over the nature of 'incomplete' or 'unsaturated' entities. More to the point, the infinite does not manifest itself in ontological problems concerning entities which somehow embody it, but in logical problems concerning the nature and status of assertions of generality.

These considerations, it seems to me, would be decisive in favour of the amended version of the theory, if it were not for the fact that there seems to be no way to account for the recursion function using these methods. There we appear actually to require Fregean functions as arguments. But let's look at the matter more closely.

Why do we need recursion in the first place? Let's consider the Cantorian version of the theory to begin with. Certainly definition by transfinite recursion is a standard technique with fairly widespread applications. In mathematical analysis, for example, one uses transfinite recursion on the countable ordinals to define the class of Baire functions. This definition, however, can be carried out without using the recursion function. In general, whenever we can confine the values of the function being defined to lie in a set which can be specified in advance, then the definition can be carried out without employing the recursion function. Such ***limited***, or ***local***, recursive definitions can be reduced to explicit ones using the familiar technique devised by Dedekind.

---

[1] It would be more accurate to say that we give a kind of substitutional *justification* of the (derived) universal generalization rule. The *meaning* of a universally qualified sentence is fixed by its definition.

Not all recursions are local ones, however. For example, we need the full force of the recursion function to show that every well-ordered set is order isomorphic to a von Neumann ordinal, or to define Cantor's $\aleph$ function. More seriously still, we need unrestricted recursion in order to establish the basic facts about the global structure of the species of all objects as a cumulative hierarchy. Indeed, without the recursion function we cannot even prove that every set has a transitive closure. The transitive closure function, $\mathrm{Tr}(\xi)$, satisfies the conditions

$$\vdash a \in \mathrm{Tr}(a) \wedge (\forall x \in \mathrm{Tr}(a))\,[x \subseteq \mathrm{Tr}(a)]$$

$$\vdash (\forall y \in P(\mathrm{Tr}(a)))\,[[a \in y \wedge (\forall x \in y)\,[x \subseteq y]] \supset \mathrm{Tr}(a) = y]$$

The transitive closure of a set is thus the smallest transitive set containing the given set as an element, where by a transitive set I mean a set for which all members of its members are also its members. This function is easily defined using the recursion function,

Let us define the ***ultimate constituents*** of a to be the set of individuals which lie in the transitive closure of a. Formally we must introduce a new function sign '$\mathrm{Con}(\xi)$' by means of the formal definition

$$\vdash \mathrm{Con}(a) = \{x \in \mathrm{Tr}(a)\colon \neg\, \mathrm{Set}(x)\}$$

Then we may regard every set as having been "built up" from its ultimate constituents in successive stages indexed by ordinal numbers, the sets formed at a given stage being composed of elements obtained at previous stages. The number of stages required to construct a set is its ***rank***. The species of all objects which are either ultimate constituents of a given set, or are sets built up from those constituents in a fewer number of stages than are required for the construction of the given set, is called the ***local universe*** of the given set. The local universe of a set a is always itself a set, $V(a)$[1]. Now this rather fanciful 'operationalistic' description of the local universe function can be replaced by a rigorous definition using the recursion function. The function $V(\xi)$ so introduced can be characterised by the following conditions:

---

[1] Strictly speaking, of course, we should not confuse the set V(a) with the species of all its members.

$$\vdash \mathrm{Set}\,(V(a))$$

$$\vdash \mathrm{Tr}(a) \subseteq V(a) \wedge (\forall x \in V(a))\,[V(x) \cup P(V(x)) \subseteq V(a)]$$

$$\vdash (\forall y \in P(V(a)))\,[[\mathrm{Tr}(a) \subseteq y \wedge (\forall x \in y)\,[V(x) \cup P(V(x)) \subseteq y] \supset y = V(a)]$$

As with the transitive closure function, there is no way to define the local universe function without using unlimited recursion. Having defined this function, however, we can then define the rank function, rang (ξ), by means of the formal definition

$$\vdash \mathrm{rank}(a) = \{x \in V(a) \colon \mathrm{Ord}(x)\}$$

From this definition we can establish that

$$\vdash \mathrm{Ord}(\mathrm{rank}(a))$$

Moreover, the local universe of a set is determined by its rank and the set of its ultimate constituents, since

$$\vdash V(a) = V(\mathrm{rank}(a) \cup \mathrm{Con}(a))$$

All of these results were first established by Zermelo in his classic paper '*Über Grenzzahlen und Mengenbereiche*' [1930].

Thus without the aid of the recursion function, we cannot establish the fundamental global structural properties of the universe of sets in its Cantorian version. Similar difficulties arise for the non-Cantorian version of the theory, though there the foundation significance of these difficulties is somewhat harder to evaluate. Without the recursion function it is not at all obvious, for example, in what sense there can be said even to be a natural number sequence. It is not possible, for instance, to prove that every finite set can be put into one-to-one correspondence with a finite von Neumann ordinal.[1] And there appear to be infinite sequences (all of whose initial segments are finite) which have different 'lengths' in some sense. Such features may appear to be pathological, but I am not at all convinced that this is the case, and I shall return to discuss the matter in due course.

So much for the practical importance of global recursion. The question we must now face is whether there is any way of accounting for global recursion if we treat function signs as syncategorematic signs. I think it very unlikely that this can be done. The form of the Axiom of Recursion differs significantly from that shared by the axioms treating the other function signs. Those axioms, you will recall, have a strong definitional air about them. They can all be regarded as stipulating membership conditions for the references of terms constructed from the function signs in question. Each has the form of an equivalence in which the defined sign occurs only on the left. That sign is therefore eliminable, not to be sure, in all contexts, but in contexts of the form '$s \in t$' where the sign in question is the principal function sign of the term "t". In the Axiom of Recursion, however, the sign '$\mathrm{Rec}_x(f(x), \xi, \zeta)$' occurs in both the '*definiendum*' and the '*definiens*'. The axiom cannot, therefore, be regarded as a mere stipulation of reference.

---

[1] This is easily proved in the Cantorian version, however, even without the Axiom of Recursion.

Like the other second level functions of our theory, the comprehension and replacement functions, the recursion function satisfies Brouwer's continuity principle. But the determination of the 'modulus of continuity' function is a much more problematic affair with recursion than with the other functions. In the case of the replacement function, for example, the modulus of continuity at arguments $\mathrm{f}(\xi)$ and a is obtained simply by applying the identity function to a. Thus if $\mathrm{f}(\xi)$ and $\mathrm{g}(\xi)$ agree at all members of a, then $\{f(x): x \in \xi\}$ takes the same value at $\mathrm{f}(\xi)$ and a as at $\mathrm{g}(\xi)$ and a. For recursion, however, the modulus of continuity at $\mathrm{f}(\xi)$, r, and a, is obtained by applying the function

$$\{(\{(z, \mathrm{Rec}_x(f(x), \xi, z)): z \in \mathrm{Seg}\,(\xi, y)\}, \xi, y): y \in \mathrm{Seg}\,(\xi, \zeta)\}$$

to the arguments $\mathrm{f}(\xi)$, r (for ξ), and a (for ζ). For if $\mathrm{f}(\xi)$ and $\mathrm{g}(\xi)$ agree at all the members of the resulting value of this function, then $\mathrm{Rec}_x(f(x), \xi, \zeta)$ takes the same value for the argument $\mathrm{f}(\xi)$, r, a as for the arguments $\mathrm{g}(\xi)$, r, a. But there seems to be no way of defining the modulus of continuity of $\mathrm{Rec}_x(f(x), \xi, \zeta)$ at $\mathrm{f}(\xi)$, r, a, without applying recursion to $\mathrm{f}(\xi)$. That modulus cannot be given in advance, as it were. It is as if we were forced to look ahead in order to anticipate values of $\mathrm{f}(\xi)$ at arguments not yet defined, and which can only be defined by recursion on $\mathrm{f}(\xi)$ itself. There is a kind of Catch 22 in operation here: we don't have to look at all of $\mathrm{f}(\xi)$ to evaluate $\mathrm{Rec}_x(\mathrm{f}(x), r, a)$; but we cannot exclude any of $\mathrm{f}(\xi)$ until we have calculated, or all but calculated, the value of the function at the given arguments.

Of course it is not really legitimate to discuss this issue in such operationalistic language. But what alternative do we have? Here we must face the fact that the intuition which tells us that recursive definition is permissable is itself an operationalistic one. The informal argument to establish that recursion is well-defined is an induction along the well-founded relation on which the recursion is based. But the induction must be with respect to a property which cannot be expressed in our conceptual notation, as it requires an unbounded existential quantifier for its formulation.

Now it is not difficult to establish the formal validity of induction along a well-founded relation in the new *Begriffsschrift*. But the form which the proof takes is of fundamental importance to the issue under discussion here. The proof of the formal proposition

$$\vdash Wf(a) \supset [(\forall x \in \text{Field}\,(a))\,[(\forall y \in \text{Seg}\,(x, a))\,[Q(y)] \supset Q(x)]$$
$$\supset (\forall x \in \text{Field}\,(a))\,[Q(x)]]$$

is accomplished by forming the set $\{x \in \text{Field}\,(a)\colon \neg\, Q(x)\}$ and then establishing, on the assumption that a is well-founded, that if the hypothesis of the induction holds, then the set $\{x \in \text{Field}\,(a)\colon \neg\, Q(x)\}$ cannot have a minimal member and must therefore be empty. But it is essential to the argument that we can form the *set* $\{x \in \text{Field}\,(a)\colon \neg\, Q(x)\}$, and this, in turn, depends crucially on the assumption that $Q(\xi)$ is a classical concept which is determinately true or false of any object. If $Q(\xi)$ is not a classical property—a *definite Eigenschaft*, in Zermelo's terminology—then the elements of the field of a which fail to have $Q(\xi)$ only form a sub*species* of the field of a which need not coincide with any of its sub*sets*. The proof thus breaks down if $Q(\xi)$ is not a classical concept.[1]

Notice that the proof breaks down even if we make the additional assumption that the well-founded relation in question is finite. Indeed, it is a familiar fact of intuitionistic arithmetic that even finite sets of natural numbers do not satisfy the classical least number principle. Since the classical proof fails so disastrously, the proof of the Principle of Induction for non-Classical properties must take the form of an iteration, along the well-founded relation in question, of a series of particular proofs that the property holds at each particular element in the field of the relation. The construction of such a 'proof' would have to take the form of a function assigning to each element x in the field of the relation a 'local proof', π_x, of the particular instance of the conclusion determined by x. And each such local proof must rest on the previous ones, where "previous" is understood in the sense of the given well-founded relation. To be sure, one is tempted to say, in the case of a finite relation, that such a construction can be carried out *in principle*. But we must ask ourselves *what* principle is involved here. And we shall see that it is the principle that recursive definitions of functions along well-founded relations are always efficacious: precisely the principle we had hoped to establish by the very sort of induction we are trying to justify.

---

[1] This will, of course, be the case if $Q(\xi)$ requires an unrestricted existential quantifier for its expression.

It looks as though we must simply take the recursion function as a primitive 'construction' or 'operation' of set theory, if that is, we are to accept it at all. Apparently we cannot, on pain of a vicious circle in our reasoning, hope to give a syncategorematic interpretation to the function sign '$\mathrm{Rec}_x(f(x), \xi, \zeta)$' by employing an induction argument to establish that every term of the form '$\mathrm{Rec}_x(f(x), a, b)$' has well-defined reference. Indeed, given the particular manner in which each value of the recursion function depends upon the values of its function argument, we seem to be driven to accepting functions as real entities if we wish to employ the recursion function in our theory.

But should we employ recursion? This question is perhaps the deepest and most difficult question in the foundations of mathematics. For the Cantorian version of our theory, it seems to me that we should include recursion, if only on the conservative grounds that without it we cannot account for the actual practice of mathematicians, nor, incidently, can we give a satisfactory explanation of the global structure of the species of objects as a cumulative hierarchy. But a decent respect for tribal pieties and institutions cannot be a guiding principle when we come to consider the strictly finitary, non-Cantorian version of the theory. For in that theory we must, in any case, violate the most fundamental principle of all those which underly conventional mathematical practice: the basic Cantorian assumption that the species of natural number forms a set.

Let us take a brief look at what is at stake for the finitary theory. Without the recursion function, we do not even have a unique natural number sequence. Let me elaborate on this point. A first level function, $f(\xi)$, (however construed ontologically) for which we can establish the formal propositions

$$\vdash \mathrm{f}(a) = \mathrm{f}(b) \supset a = b$$

and

$$\vdash \neg\,[\mathrm{f}(a) = \varnothing]$$

can be taken as determining an ***infinite sequence***, namely, the sequence

$$\varnothing,\ \mathrm{f}(\varnothing),\ \mathrm{f}(\mathrm{f}(\varnothing)), \ldots$$

Call such a function, $\mathrm{f}(\xi)$, a ***scale***, if for any set (which must be finite in the present case), there is[1] a finite initial segment of the sequence generated by f(ξ) which can be placed in one-to-one correspondence with the given set. In short, the function f(ξ), or the infinite sequence generated by it, is a scale if it *measures* every finite set. With the recursion function we can easily show that the successor function, $\xi \cup \{\xi\}$, for the von Neumann natural numbers is a scale. Without recursion there is no reason to believe that any function is a scale. We cannot even show that the von Neumann natural numbers are closed under addition (*i.e.* that if a and b are both measured by von Neumann natural numbers, then so is $a \cup b$). This is perhaps not surprising given the extremely slow rate of growth of the von Neumann ordinals. We can, however, easily define infinite sequences which extend[2] a given infinite sequence and are closed under both addition and multiplication.[3] But there is no obvious way to modify these methods to define an infinite sequence extending a given sequence and closed under exponentiation.[4] Nevertheless, the 'arithmetic' of this general theory of finite sets does clearly include exponentiation in some sense, for the presence of the power set function allows us to define functions which enlarge the cardinality of their arguments at exponential rates. The 'natural number arithmetic' which describes the facts of cardinality embodied in this theory of finite sets is the free variable arithmetic based on Grzegorczyk's class $\mathscr{E}_3$ of elementary functions.

The view of arithmetic which emerges from this minimal version of finitary set theory without a recursion function is a strange and unfamiliar one. But may it not, perhaps, be the correct view? Curiously enough, our new *Begriffsschrift*, in this extreme finitary form, has features which resemble the ultra-intuitionism of Essenin–Volpin. I say 'curiously' because these two theories are polar opposites in their approaches to finitary foundations. Essenin–Volpin's theory is an operationalist theory based upon the notion of the 'feasibility' of carrying out actual calculations with numbers. But the new *Begriffsschrift*, even in its strictly finitary version, remains severely classical in outlook. The guiding principle of the theory, like that of its namesake, is to avoid the slightest hint of dependence on intuitive (*anschaulich*) notions like 'operation' or 'iteration'. It embodies the classical view of Dedekind and Frege that such notions have no place in the foundations of mathematics.

---

[1] We can avoid the use of unlimited quantification in this definition of *scale* if we rephrase it to say that a definite function sign, g(ξ), can be *exhibited* which satisfies the appropriate formal propositions.

[2] The sequence generated by g(ξ) *extends* that generated by f(ξ) if there is a function h(ξ) which when applied to an initial segment of the f(ξ) sequence yields an initial segment of the g(ξ) sequence having the same cardinality.

[3] A sequence is closed under multiplication if it measures $a \times b$ whenever it measures both a and b.

[4] A sequence is closed under exponentiation if it measures a^b whenever it measures both a and b. a^b is the set of all local functions defined on b with values in a.

It is a widely held view that the question 'What is a natural number?' is irrelevant to mathematics, however important it may be for philosophy. For Cantorian mathematics it clearly *is* irrelevant. But for strictly finitary mathematics it may be of decisive importance. For if we take the view that 'the natural number sequence' is simply a datum, that it is the business of finitary foundations to explore its structure in accordance with whatever canons of argument and definition are deemed to be finitarily acceptable, then we may render ourselves blind to the really important issues, for example, the problem of analysing recursive definitions. And, incidently, we may find it exceedingly difficult to free our imaginations from the grip of the traditional 'classical' picture of the natural numbers. For at the root of the traditional picture of the natural numbers is the conviction, which can, of course, be justified by proofs if we make Cantorian assumptions, that there are uniform methods for selecting a unique representatives of every finite cardinality. The traditional natural number sequence is an abstract representation of the result of such a uniform selection. But it may be precisely this conviction which must be called into question in finitary foundations.

The new *Begriffsschrift*, in both its finitary and Cantorian guises, takes, as the fundamental notion underlying mathematics, the primitive notion of number as a quantitatively determined plurality or set, a notion which is to be found in the very origins of our mathematical culture in classical antiquity. The more recent abstract concept of number constructed from the original concept is dismantled, although concrete versions, or partial versions, of it can be reassembled from more primitive elements.

We began our discussion in this section by inquiring into the ontological status of functions and concepts. Where, then, has our inquiry led us? Of all the possibilities which have emerged from our discussion, there are, it seems to me, two which stand out as the most coherent. The first of these is the Cantorian version of the theory based on an ontology of objects, functions, and concepts. This theory is a conservative one. It makes explicit the implicit assumptions underlying modern mathematics. For this reason it holds no surprises for us. The proof procedures described in section 4(*i*) are necessarily incomplete. But all known techniques of mathematical proof can be justified by these procedures.[1] Similarly, the theory of definition embodied in the system suffices to carry out all known mathematical definitions. Indeed we may perhaps claim for our theory of definition a completeness to which our theory of proof cannot possibly aspire. Our theory is thus completely general in the following sense: any given mathematical theory can be expounded in a suitable definitional extension of our basic conceptual notation. Of course our theory does not contain all of mathematics, because it does not contain all mathematical definitions. But it does constitute a logical framework for all of mathematics.

---

[1] This follows from the fact that the system is a 'free variable' version of Zermelo–Fraenkel set theory. It can be shown that the first order part of the system, which includes the logic of the quantifiers $(\forall x \in t)$ and $(\exists x \in t)$, is complete.

The second version of the theory which recommends itself to us on the basis of its foundational coherence is the strictly finitary version based on the single ontological category of objects. Treating incomplete signs as syncategoremata, rather than as names of functions or concepts, seems more in the spirit of strict finitism. And it carries with it the consequence that definition by recursion is not a primitive notion of the theory. It would seem appropriate, however, to include the transitive closure function $\mathrm{Tr}(\xi)$, among the primitive notions of the theory. To be sure, this raises difficulties for our syncategorematic reading of function signs. For I can see no way of characterising the transitive closure which puts it in the form of a 'formal stipulation' of the sort used in introducing the other function signs. The simplest course of action seems to be simply to lay down the following new axioms:

VIII. Axioms characterising the Transitive Closure of a Set.

1. $\vdash a \in \mathrm{Tr}(a) \wedge (\forall x \in \mathrm{Tr}(a))\,[x \subseteq \mathrm{Tr}(a)]$
2. $\vdash (\forall y \in P(\mathrm{Tr}(a)))\,[[a \in y \wedge (\forall x \in y)\,[x \subseteq y]] \supset y = \mathrm{Tr}(a)]$

But the assumption that the species of elements of the transitive closure of a set themselves form a set, seems to me to be a very natural 'finiteness assumption', analogous, say, to the assumption, incorporated in our theory by the formal stipulation introducing the power set function, $P(\xi)$, that the species of all subsets of a set forms a set. For the axioms of VIII say that the elements which enter into the 'construction' of a set must be definite in number.

Naturally the set constant ω must be dropped, and a final axiom to the effect that all sets are finite must be adopted. Like the Cantorian version of the theory, the strictly finitary version is open to indefinite extension via definitions, and the addition of new basic laws and rules of proofs.

5 CONCLUSIONS

The new *Begriffsschrift*, in its technical details, derives more from Zermelo than from Frege. Moreover, some of the most important central ideas underlying it have been taken from Cantor and from Brouwer. In what sense, then, can it legitimately be described as a new version of Frege's theory? There are, I believe, three important respects in which it can only be described as a Fregean theory.

The first of these is that the theory as a whole rests upon a general system of metaphysics, moreover, a system very like the system Frege himself developed. It was one of Frege's greatest insights that the roots of mathematical definition and proof must lie in a simple, general ontology. None of the other thinkers whose ideas inform our theory saw this so clearly, if indeed, they saw it at all. Zermelo, at least in his earliest work [1908], believed he was setting up an axiomatic theory of sets analogous to Dedekind's axiomatic theory of real numbers. He did not, at that time, notice that set theory, of all theories, cannot be presented axiomatically because of its central role in the explication of axiomatic definition. Cantor's metaphysics was not as carefully thought out as Frege's, nor as scrupulously presented. Moreover, as we have seen, he placed great store in the Aristotelean doctrine of abstraction, and this led him into the errors of psychologism. Cantor did not separate, as Frege did, logical and metaphysical questions from psychological and epistemological ones. Brouwer's outlook was, of course, self-consciously psychologistic. And to this day, his followers, the intuitionists, continue to inhabit a kind of pre-Fregean philosophical arcadia, where they are content to identify the objects of mathematics with 'mental constructions', without, it would seem, ever troubling themselves with the question of just what such things might possibly be. All of this is in strong contrast with Frege's single-minded and persistent pursuit of objectivity.

The second respect in which our theory is a Fregean one is to be found in the fact that it is a general, universal system of classical logic. Moreover, its primitive notions are not subject to varying interpretations, in the manner of an axiomatic theory, but have fixed, determinate meanings: this is also a Fregean characteristic. Indeed it is possible, although I have not emphasised this here, to regard standard predicate logic as a fragment of the new *Begriffsschrift*, which itself can be regarded as a universal higher order logic of absolutely infinite order. And viewing the matter in this way, allows us to solve certain perplexing difficulties in the foundations of mathematics, difficulties arising out of a failure to grasp the essential connection between set theory and classical logic. I have discussed this point in detail in my article 'On the Consistency Problem for Set Theory' ([1977], see especially Part I), and so I shall not repeat myself here.

Finally Frege's fundamental distinction between objects and functions is of central importance for the exposition of the new *Begriffsschrift*. It bears upon virtually all of the important issues which must be taken into account for that exposition. Indeed, I have not been able to resolve the mathematical and metaphysical difficulties surrounding this distinction, although I hope that I have been somewhat successful in drawing attention to them. In this distinction it seems to me we have Frege's most original contribution to the foundations of mathematics. One must acknowledge this if one considers the connection between this doctrine and Frege's logical grammar. The function-object distinction, and its grammatical correlate underlies the whole of modern mathematical logic.

University of Bristol

REFERENCES

ARISTOTLE [1934]: *The Physics* (2 vols), translated by P. H. Wicksteed and F. M. Cornford, The Loeb Classical Library, London: William Heinemann Ltd. [1956], *The Metaphysics*, translated by John Warrington, Everyman's Library No. 1000, London: J. M. Deut and Sons Ltd.

CANTOR, G. [1897]: *Contributions to the Founding of the Theory of Transfinite Numbers*, translated by P. E. B. Jourdian, New York: Dover.

[1932], *Gesammelte Abhandlungen*, E. Zermelo (*ed.*), Hildesheim: Georg Olmo Verlagsbuchhandlung.

CHURCH, A. [1956]: *Introduction to Mathematical Logic*, Princeton: Princeton University Press.

DUMMETT, M. [1973]: *Frege, Philosophy of Language*, London, Duckworth.

FREGE, G. [1879]: *Conceptual Notation*, translated by T. W. Bynum, Oxford: Clarendon Press.

[1884]: *The Foundations of Arithmetic*, translated by J. L. Austin, Oxford: Basil Blackwell.

[1892]: 'Concept and Object' in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (*eds.*), Oxford: Basil Blackwell.

[1892*a*]: 'Sense and Reference' in P. Geach and N. Black (*eds.*): *Translations From the Philosophical Writings of Gottlob Frege*, Oxford: Basil Blackwell.

[1893]: *The Basic Laws of Arithmetic*, translated and edited by M. Furth, Berkeley and Los Angeles: University of California Press.

[1894]: Review of E. Husserl *Philosophie der Arithmetik*, in P. Geach and M. Black (*eds.*): *Translation from the Philosophical Writings of Gottlob Frege*, Oxford: Basil Blackwell.

GÖDEL, K. [1947]: 'What is Cantor's Continuum Problem?' *American Mathematical Monthly*, **54**, pp. 125–53.

KLEIN, J. [1968]: *Greek Mathematical Thought and the Origins of Algebra*, London: M.I.T. Press.

KREISEL, G. [1967]: 'Informal Rigour and Completeness Proofs' in I. Lakatos (*ed.*): *Problems in the Philosophy of Mathematics*, Amsterdam: North Holland, [1969]: 'Mathematical Logic' in T. L. Saaty (*ed.*): *Lectures on Modern Mathematics* (vol. III), New York: John Wiley and Sons.

KREISEL, G. and KRIVINE, J. L. [1967]: *Mathematical Logic*, Amsterdam: North Holland.

MAYBERRY, J. P. [1977]): 'On the Consistency Problem for Set Theory, An Essay on the Cantorian Foundations of Classical Mathematics', *British Journal for the Philosophy of Science*, **28**, pp. 137–70.

SCHOPENHAUER, A. [1958]: *The World as Will and Representation*, translated by E. J. F. Payne, New York: Dover.

ZERMELO, E. [1908]: 'Investigations in the Foundations of Set Theory', translated by S. Bauer-Mengelberg, in J. van Heijenoort (*ed.*): *From Frege to Gödel*, Cambridge: Harvard University Press.

ZERMELO, E. [1930], 'Uber Grenzzahlen und Mengenbereiche', *Fundamenta Mathematica*, **16**, pp. 29–47.

Brit. J. Phil. Sci. 31 (1980), 359–384 *Printed in Great Britain*

Review Articles

OBJECTIVITY, RATIONALITY, INCOMMENSURABILITY, AND MORE*

Few readers of this journal will need to be reminded of the provocative view of science provided by Thomas S. Kuhn. Since the publication of his *The Structure of Scientific Revolutions*,[1] Kuhn's work has generated an enormous amount of discussion and controversy. Perhaps the most vehement discussion has focused on Kuhn's conception of cross-paradigm debate, for this conception, it is argued, portrays the process of theory choice as irrational and subjective, and so threatens the philosophical ideals of rationality and objectivity. In his book *The Essential Tension* Kuhn attempts, among other things, to extend and clarify his position *vis-à-vis* the interrelated issues of objectivity, rationality, incommensurability, and relativism. While only a small number of the papers appearing in *The Essential Tension* speak directly to these issues, because of their importance for contemporary debate in philosophy of science I shall focus the bulk of this essay on them. I shall attempt to fit Kuhn's discussion of these issues into the rest of the relevant literature, including other recent papers by Kuhn not included in the volume under review. The effect will be to make the present essay not only a review of that volume, but in addition a general account of the progress and status of the discussion Kuhn's work has spawned. First, however, a general review of Kuhn's book is in order.

1 GENERAL OVERVIEW

The Essential Tension is a collection of papers, all but two previously published. A lengthy but important preface describes the slow process through which Kuhn developed his view of scientific revolutions and paradigms. The rest of the papers are divided into two sections: Part One, titled 'Historiographic Studies'; and Part Two, titled 'Metahistorical Studies'. In Part One, the first paper (one of the two previously unpublished papers in the volume), 'The Relations Between the History and the Philosophy of Science', urges philosophers of science to take seriously 'the fundamental conceptual readjustment required of the historian to recapture the past or, conversely, of the past to develop toward the present' (xiv).[2] This particularly Kuhnian emphasis on the necessity for 'conceptual readjustment' to come to grips with the relationship between the science of the past and of the present is filled out, by explication and by example, in several of the remaining papers. The second paper, 'Concepts of Cause in the Development of Physics', outlines changes in the conception, and relative importance, of that notion in physical science from Aristotle to the present; and in the relationship between the notion of cause and that of explanation or explanatory power. The next two essays, 'Mathematical versus Experimental Traditions in the Development of Physical Science' and 'Energy Conservation as an Example of Simultaneous Discovery', provide fascinating discussions of the topics their titles suggest, and also serve to illustrate Kuhn's theses (developed in the fifth essay) concerning the relative power of internal versus external factors in the development of physics. The fifth essay, titled 'The History of Science', provides a general discussion of that discipline, with specific attention being paid to the internal/external distinction and its importance for the history of science. The sixth essay, 'The Relations between History and the History of Science', is a companion-piece to the first and fifth essays, and together these papers provide a clear and unified account of the relations between the various academic disciplines (not counting the physical sciences)—philosophy of science, history, and history of science—that Kuhn's work draws upon.

---

* Review of T. S. Kuhn [1977]: *The Essential Tension*. Chicago: University of Chicago Press. £12.95. Pp. xxiii + 366.

[1] Kuhn [1970*a*].

[2] In what follows parenthetical references to page numbers should be understood as references to *The Essential Tension*. References to other works will be so labelled.

The essays in Part Two move from the historical to the metahistorical, and in several cases allow the reader to follow Kuhn's intellectual biography, as the theses formulated in *The Structure of Scientific Revolutions* evolve and develop in Kuhn's thinking (*cf.* pp. xvi–xxiii). The seventh and eighth essays, 'The Historical Structure of Scientific Discovery' and 'The Function of Measurement in Modern Physical Science', display early formulations of now familiar Kuhnian theses concerning the nature of scientific revolutions and of normal science. The ninth, 'The Essential Tension: Tradition and Innovation in Scientific Research', gives the collection of essays its title, and puts further flesh on Kuhn's conception of the interplay between normal and revolutionary science.[1] The tenth essay, 'A Function for Thought Experiments', discusses the role of thought experiments in theorising, while the fourteenth and final essay, 'Comment on the Relations of Science and Art', treats of the relations and differences between the two fields, especially noting 'the ways in which differences in shared values ... may decisively affect the developmental patterns characteristic of science and art' (p. xxi).

I have skipped over essays eleven through thirteen, for reasons I shall explain shortly. First, however, I should like to say that the essays thus far enumerated, while they will not play any further role in this review, are by and large full of insight, and demonstrate fully the importance of Kuhn's work. The historiographical essays in particular are rich and, occasionally, provocative, and offer a unique perspective both on specific episodes in the history of science and on the proper approach to the study of that subject. Whatever one thinks of Kuhn's philosophy of science, one cannot deny his importance as a first-rate historian of science. The excellent work begun with his *The Copernican Revolution*[2] continues in these essays.

The three essays I have yet to mention speak directly to the issues of objectivity, rationality, relativism and incommensurability to which the rest of this review will be devoted. The eleventh, 'Logic of Discovery of Psychology of Research', is the well-known paper in which Kuhn contrasts his views with those of Karl Popper. The twelfth, 'Second Thoughts on Paradigms', attempts to clarify and improve Kuhn's early presentations of the concept of paradigm. The thirteenth, 'Objectivity, Value Judgment, and Theory Choice' (the other of the two previously unpublished papers appearing in *The Essential Tension*) 'aims to answer the charge that I [Kuhn] make theory choice entirely subjective' (p. xx). This last essay, especially, will be treated in detail in what follows, because it is previously unpublished and because it speaks to the most controversial of Kuhn's philosophical views. First, however, it is necessary to set the stage for its consideration by reviewing in some detail the early stages of debate, between Kuhn and his critics, over the interrelated issues mentioned earlier. I hope to show that Kuhn's more recent remarks, especially his discussion in 'Objectivity, Value Judgment, and Theory Choice', do little to clarify or make plausible his original treatment of these issues.

---

[1] The interested reader may wish to compare Kuhn's discussion of the tension between tradition and innovation with W. V. Quine's view of the tension between conservatism and the striving for simplicity characteristic, in his view, of scientific research and theorizing. *Cf.* Quine [1960], pp. 20–21.

[2] Kuhn [1957].

2 THE FIRST CRITICISMS OF THE 'IRRATIONALITY THESIS'

It will be recalled that those aspects of Kuhn's early view which suggest that paradigm-choice is irrational (henceforth the 'irrationality thesis') stem largely from his contention that paradigms are incommensurable. Incommensurability—the inability of competing paradigms to be directly compared, or judged according to a neutral standard—stems from Kuhn's contention that a paradigm contains its own criteria of evaluation, as well as laws and methods of application. During paradigm debate, according to Kuhn, competing paradigms are evaluated according to paradigm-bound criteria of evaluation; because of incommensurability, there are no paradigm-neutral criteria of evaluation. Since there are no paradigm-neutral criteria of evaluation, paradigm debate can rely on no objective criteria of evaluation of paradigms; hence paradigm debate is irrational. The irrationality thesis thus rests on incommensurability, which in turn rests on the paradigm-bound nature of criteria of evaluation of paradigms. The irrationality thesis is therefore open to criticism through criticism of the paradigm-bound conception of criteria of evaluation of paradigms.[1] This is, in fact, a major focus of some of the critics of the irrationality thesis. Scheffler, Shapere and Kordig, for example, have attacked the irrationality thesis in just this way—by criticising Kuhn's arguments for the paradigm-bound nature of criteria of evaluation of paradigms. Kuhn, it will be recalled, defines paradigms as including methods and criteria of evaluation as well as theory, so that proponents of competing paradigms assess paradigms according to different, incompatible sets of criteria of evaluation of paradigms. As Scheffler describes Kuhn's position:

---

[1] Other criticisms focus on Kuhn's derivation of incommensurability from his views on *meaning-variance*—*i.e.*, his view that scientific terms change meaning from theory to theory. I shall not discuss this aspect of Kuhn's view in this paper, for it has been dealt with at length and, in my opinion, convincingly refuted by Dudley Shapere (Shapere [1964], [1966]), Israel Scheffler (Scheffler [1967], chapter 3) and Carl R. Kordig (Kordig [1971], chapters 2 and 3). Kuhn has also presented incommensurability in terms of 'radical translation' (Kuhn [1969], pp. 200–4; [1970*b*], pp. 267–77; [1971], p. 146; and [1977], pp. xxii–xxiii and 338–9). While I shall not discuss this at present, I believe that Kuhn's conceiving of scientists of competing paradigms as being engaged in radical translation is confused, and crucially depends on his confused discussions of meaning variance just mentioned. For brief comments on Kuhn's conception of radical translation and its relevance to incommensurability, see Shapere [1971], p. 708. For more recent discussion of meaning change, theory change, and incommensurability, all of which cast serious doubt on the Kuhnian position, see Kitcher [1978]; Moberg [1979]; and Levin [1979].

> . . . to accept a paradigm [according to Kuhn] is to accept not only theory and methods, but also governing standards or criteria which serve to justify the paradigm as against its rivals, in the eyes of its proponents. Paradigm differences are thus inevitably reflected upward, in criterial differences at the second level. It follows that each paradigm is, in effect, inevitably self-justifying, and that paradigm debate must fail of objectivity.[1]

Scheffler criticises this view of Kuhn's by arguing that it blurs a crucial distinction by failing to distinguish between criteria *internal* to a paradigm and criteria *external* to a paradigm. Kuhn is perhaps correct that each paradigm sets criteria such that within a paradigm, certain pieces of evidence or experimentation can only be interpreted or evaluated according to the criteria set forth by the paradigm; however, he has failed to produce any plausible reason for thinking that external criteria—this is, criteria according to which paradigms are themselves judged—must be included in a paradigm. Paradigm debate is more properly construed as a second-order, or meta-, activity, and at this higher level of evaluation independent criteria are operative. The distinction between internal and external criteria thus renders gratuitous Kuhn's claim that paradigms are 'self-justifying'—for in light of the internal/external distinction, it must be concluded that competing paradigms are judged according to external criteria, hence not according to criteria which stem from and thus favour one of the paradigms being judged as against its rivals. Kordig, in agreement with Scheffler, writes that:

> Kuhn has failed to show that paradigm differences imply evaluative differences at a second, or meta, level; he has not shown that the sharing of second-order criteria by enthusiasts of rival paradigms is impossible. He has, therefore, not demonstrated that scientific transitions consist only in non-cumulative persuasions and conversions.[2]

By bringing to light this crucial objection to Kuhn's argument for the paradigm-bound nature of criteria of evaluation of paradigms, Scheffler and Kordig have cast great doubt on Kuhn's incommensurability claim—for if second order principles are operative in paradigm-evaluation, there need be no incommensurability between competing paradigms, since they are capable of being evaluated according to these second-order principles. Denial of incommensurability, in turn, renders suspect the irrationality thesis, since that thesis, as we saw earlier, depends in large part on incommensurability.

---

[1] Scheffler [1967], p. 84.

[2] Kordig [1971], p. 106.

3 KUHN'S INITIAL RESPONSE

Kuhn initially responded to these criticisms not by defending the irrationality thesis, but by denying the irrationality thesis and complaining that his critics misinterpret his position.[1] He writes:

> My critics respond to my views . . . with charges of irrationality, relativism, and the defense of mob rule. These are all labels which I categorically reject.[2]

Kuhn claims that his critics err when they report that he believes that:

> the proponents of incommensurable theories cannot communicate with each other at all; as a result, in a debate over theory-choice there can be no recourse to *good* reasons; instead theory must be chosen for reasons that are ultimately personal and subjective; some sort of mystical apperception is responsible for the decision actually reached.[3]

If this is misinterpretation, what is the correct interpretation of Kuhn's views? Kuhn clarifies his position by claiming, if effect, that there is no algorithm, no rules, for theory-choice. He argues that there are never deductive proofs which necessitate the choice of one theory over another, and that debates over theory-choice can never end with a conclusive logical verdict.[4] This is, as Kuhn recognises, hardly an unfamiliar thesis; indeed, it is a proposition generally agreed to by all parties in the debate. Why, then, if this is all Kuhn meant to say concerning paradigm-choice, all the rhetoric about 'conversion', 'gestalt switches', and the like during scientific revolutions? Kuhn attempts to clarify his position through the introduction of 'values':

> Nothing about that relatively familiar thesis [*i.e.*, that there is no algorithm for theory choice] implies either that there are no *good reasons* for being persuaded or that these reasons are not ultimately decisive for the group. Nor does it even imply that the reasons for choice are different from those usually listed by philosophers of science: accuracy, simplicity, fruitfulness, and the like. What it should suggest, however, is that such reasons function as *values* and that they can thus be differently applied, individually and collectively, by men who concur in honoring them. If two men disagree, for example, about the relative fruitfulness of their theories, or if they agree about that but disagree about the relative importance of fruitfulness and say, scope in reaching a choice, neither can be convicted of a mistake. Nor is either being unscientific. There is no neutral algorithm for theory-choice, no systematic decision procedure which, properly applied, must lead each individual in the group to the same decision.[5]

---

[1] Kuhn's earliest published replies to his critics are his Kuhn [1969], and his [1970*b*]. The present section of this paper, dealing with Kuhn's initial responses to his critics, will focus solely on these two essays. Later replies to his critics will be discussed in subsequent sections.

[2] Kuhn [1970*b*], p. 234.

[3] Kuhn [1969], pp. 198–9, emphasis Kuhn's. Kuhn's later replies echo these earlier ones. See, for example, Kuhn [1977], pp. 336–9, esp. p. 337.

[4] Kuhn [1969], p. 199. For further discussion of this point *vis-à-vis* Kuhn and his critics, see Scheffler [1972], p. 369.

[5] Kuhn [1969], pp. 199–200, emphasis added.

This paragraph of Kuhn's seems unexceptionable (disregarding for a moment the mention of 'values'), and is a far cry from the irrationality thesis. There is no mention here of paradigm-bound criteria of evaluation; no mention of incommensurability. However, there is a question about the notion of 'values' Kuhn introduces, for the relationship between values and the irrationality thesis needs to be spelled out. It also needs to be pointed out that Kuhn, in this passage, allows himself recourse to *good reasons* during paradigm-choice; whether or not Kuhn is entitled to do so will also be considered presently. For now, I wish to stress the non-radicalness of the cited paragraph, and the extent to which it softens the claims of Kuhn's original presentation of the irrationality thesis. Let us now turn our attention to the remainder of Kuhn's remarks. He continues the passage:

> . . . To understand why science develops as it does, one need not unravel the details of biography and personality that lead each individual to a particular choice, though that topic has vast fascination. What one must understand, however, is the manner in which a particular set of shared values interacts with the particular experiences shared by a community of specialists to ensure that most members of the group will ultimately find one set of arguments rather than another decisive.
>
> That process is persuasion, but it presents a deeper problem. Two men who perceive the same situation differently but nevertheless employ the same vocabulary in its discussion must be using words differently. They speak, that is, from what I have called incommensurable viewpoints.[1]

Here Kuhn has brought back into the discussion incommensurability, and consequently, the irrationality thesis. The tension between the two cited passages runs throughout Kuhn's initial responses, and is in large part the reason his critics fail to be quieted by these responses. He continually tries both to retract the more radical claims of his earlier writing, and at the same time to maintain them under different guises. Besides the just-cited re-introduction of incommensurability, Kuhn's newer views depend in large part on the notions of good reasons and values, notions which are not clearly reconcilable with his original presentation:

> What I am denying then is neither the existence of good reasons nor that these reasons are of the sort usually described. I am, however, insisting that such reasons constitute values to be used in making choices rather than rules of choice.[2]

These remarks make Kuhn's overall position difficult to make sense of. It is not clear that these just-cited passages concerning the role of values, and the existence of good reasons for theory choice, are consistent with his original presentation of incommensurability. Consider, for example, Kuhn's emphasis on one particular characteristic of values: they 'may be shared by men who differ in their application',[3] so that two scientists may share the value of, say, simplicity, yet differ as to the simplicity of their competing theories, or the importance of

[1] Kuhn [1969], p. 200.
[2] Kuhn [1970*b*], p. 262.
[3] Kuhn [1969], p. 185.

simplicity relative to another value, say scope. This is unexceptionable, and is little more than elaboration of the 'no algorithm for theory choice' theme. However, it is at odds with incommensurability, if Kuhn means to suggest that the two scientists can deliberate meaningfully on the question of the fruitfulness of their respective theories. If Kuhn is willing to allow for such deliberation, which some of the just-cited remarks seem to indicate, his comments on values are clearly inconsistent with his maintenance of incommensurability, for incommensurability guarantees the impossibility of such deliberation. If, on the other hand, Kuhn is not willing to allow for such deliberation, it is difficult to see how the introduction of 'values' into the discussion enlarges on, or clarifies, the incommensurability claim. The status of values, then, needs to be clarified. As Shapere puts the problem in his review of Kuhn's early replies to his critics, 'if there are . . . no constraints on what one can assert in the name of "values" ', Kuhn's viewpoint is 'as relativistic, as antirationalistic, as ever'.[1] If there are constraints on what one can assert in the name of values, on the other hand, it is difficult to see how Kuhn can reconcile those constraints with his incommensurability claim—for incommensurability demands the absence of such constraints.

This ambiguity permeates Kuhn's discussion of *good reasons* for theory choice as well. Kuhn insists that there are good reasons for theory choice, but that these reasons 'constitute values'. Because Kuhn's account of values is unclear, his account of reasons is as well. For if there are no constraints on what one can assert in the name of values, it is difficult to see the sense in which such assertions can be taken to constitute objective reasons for choosing one theory over another. Moreover, given incommensurability, it is difficult to make any sense of the term 'good reason' as it applies to paradigm debate, since, as Shapere puts it, 'the very notion of a "good reason" has been made paradigm-dependent'.[2]

Kuhn's introduction of 'values' and 'good reasons', therefore, seem unhelpful in clearing up the debate over the irrationality thesis. As we have seen, Kuhn has explicitly denied the thesis[3]; explicitly maintained incommensurability, from which the thesis follows[4]; and introduced into the discussion notions which seem to do nothing but further cloud his view. If we take Kuhn's denial of the irrationality thesis seriously, he may then be a proponent of objectivity in science—thus agreeing with his critics. However, it is difficult to take that denial seriously, in light of his reluctance to give up the incommensurability thesis, his hedging over the status of reasons, and his continued descriptions of theory choices as gestalt switches and conversion experiences:

> The conversion experience that I have likened to a gestalt switch remains, therefore, at the heart of the revolutionary process. Good reasons for choice provide motives for conversion and a climate in which it is more likely to occur. Translation may, in addition, provide points of entry for the neural programming that, however inscrutable at this time, must underlie conversion. But neither good reasons nor translation constitute conversion, and it is

---

[1] Shapere [1971], p. 708.

[2] *Ibid.*, pp. 707–8.

[3] Kuhn [1970*b*], p. 234; Kuhn [1969], pp. 198–9.

[4] Kuhn [1969], p. 200.

> that process we must explicate in order to understand an essential sort of scientific change.[1]

As the above passage indicates, Kuhn 'seems unwilling to abandon "incommensurability" while trying, unsuccessfully, to assert that communication and comparison are possible'.[2] Kuhn seems to want it both ways: he wants to maintain incommensurability (and so irrationality), yet deny irrationality and allow for communication between proponents of competing paradigms (thus giving up incommensurability). It is clear, I hope, that Kuhn cannot have it both ways. His maintenance of incommensurability vitiates his denial of the irrationality thesis.

4 KUHN'S FURTHER RESPONSE

Thus far, I have been reviewing the early debate between Kuhn and Scheffler, Shapere and Kordig over the objectivity of scientific theory choice. We have seen that Kuhn's early responses to these critics have been unconvincing, and have done little to enhance the plausibility of his original presentation of incommensurability. In this section of the present paper I shall consider three more recent attempts by Kuhn to answer his critics on the interrelated issues of objectivity, irrationality, relativism and incommensurability. I will argue that Kuhn's further responses fail either (1) to establish a tenable irrationality or incommensurability thesis; or (2) to eliminate the tension, illustrated in the previous section of this paper, between Kuhn's wanting to maintain a strong incommensurability claim on the one hand, and his wanting to weaken or eliminate such a claim, on the other hand.

(a) *Comparability*

At least part of the reason that incommensurability leads to arbitrary, and so irrational, choice among competing scientific theories is that incommensurability seems to preclude the objective comparability of rival theories. This is clearly the thrust of Kuhn's *The Structure of Scientific Revolutions*: incommensurability precludes comparability. Whether or not incomparability follows from incommensurability, though, is a matter of some debate. Scheffler, for example, has argued that incomparability does not follow from incommensurability, for to recognise incommensurable rival theories as *rival* is to recognise some common meta-domain in which they may be meaningfully compared (if not, in what sense can the theories be said to be competitive, or genuine rivals?) Consequently, Scheffler argues, even if Kuhn is right about incommensurability, nothing concerning comparability follows—and so, incommensurability poses no threat to the possibility of objective comparison of rival theories, and therefore no threat to the possibility of rational theory choice.[3] Kuhn has, in his recent 'Theory-Change As Structure-Change: Comments on the Sneed Formalism',[4] addressed this issue of the relationship between incommensurability and comparability. Unfortunately, Kuhn's discussion both fails to come to grips with Scheffler's criticisms of the original 'incommensurability precludes comparability' argument, and fails to clarify or refine his position concerning incommensurability. He begins his discussion by seemingly granting Schleffler's point that incommensurability does not preclude comparability:

---

[1] *Ibid.*, p. 204.

[2] Shapere [1971], p. 709.

[3] Scheffler [1967], chapters 2–4, esp. p. 82. Note also Scheffler's reference to Shapere here, and Kordig [1971], p. 100. L. Laudan argues similarly (Laudan [1976], p. 596) that 'incommensurability of theories at the object-level does not entail incomparability at the meta-level'.

[4] Kuhn [1976].

> Most readers of my text have supposed that when I spoke of theories as incommensurable, I meant that they could not be compared. But 'incommensurability' is a term borrowed from mathematics, and it there has no such implication. The hypotenuse of an isosceles right triangle is incommensurable with its side, but the two can be compared to any required degree of precision. What is lacking is not comparability but a unit of length in terms of which both can be measured directly and exactly[1]

These remarks, it seems, grant Scheffler's point that incommensurability does not preclude comparability—and so, that objective comparisons of rival scientific theories is possible. Taken by themselves, these remarks constitute, if not a complete repudiation, at least a substantial weakening of Kuhn's original position on incommensurability and comparability. Unfortunately, however, Kuhn continues his comments here in a way which renders his overall position confusing. He continues the just-cited passage as follows:

> 'In applying the term "incommensurability" to theories, I had intended only to insist that there was no common language within which both could be fully expressed and which could therefore be used in a point-by-point comparison between them.'[2]

This last sentence reintroduces the incomparability the earlier remarks deny. For unless there is a substantive difference between 'comparison' and 'point-by-point comparison', Kuhn is saying that incommensurable paradigms can be compared, but not compared 'point-by-point'. This is equivalent to saying that they can be compared, but not compared, which does little to illuminate Kuhn's position. Moreover, in a footnote at this juncture in his paper, Kuhn pointedly returns to his original 'incommensurability precludes comparability' position, as well as his original view that competing theories are indeed incommensurable, so that objective, rational theory choice is impossible:

> When I first made use of the term 'incommensurability', I conceived the hypothetical neutral language as one in which any theory at all might be described. Since then I have recognized that comparison requires only a language neutral with respect to the two theories at issue, but I doubt that anything of even that limited neutrality can be designed.[3]

What can be made of this? In the space of a single paragraph and footnote, Kuhn has allowed for the possibility of neutrally comparing rival scientific theories (thus weakening, if not repudiating, his original incommensurability thesis); doubted the possibility of such neutral comparison; and denied the possibility of 'point-by-point' neutral comparison (while failing to explain the alleged difference between 'point-by-point 'comparison and comparison *simpliciter*). It seems clear that these recent comments by Kuhn do little to clarify or make plausible his position concerning incommensurability, comparability, and the objectivity of scientific theory choice. As we saw in the previous section of this paper concerning Kuhn's initial response to his critics, this later response does little to eliminate the tension between Kuhn's wanting both to maintain a strong incommensurability claim, and simultaneously to allow for rational and objective theory choice in science. And, again as we saw earlier, he cannot have it both ways.

---

[1] *Ibid.*, pp. 190–1.
[2] *Ibid.*, p. 191.
[3] *Ibid.*, footnote 11, p. 198.

(b) *Relativism*

Kuhn's critics have also argued that Kuhn's incommensurability thesis rendered his view relativistic, in that acceptance of a paradigm or theory was relative to one's prior commitment to the paradigm or theory in question, thus making objective evaluation impossible. In later responses to his critics, Kuhn has attempted to refine his original position concerning relativism, and to show that he is not a relativist in any 'bad' sense:

> If you would think of science's developing in the usual treelike pattern, then I think it not only possible, but indeed quite easy, to do the following: design a set of criteria which would permit a neutral observer to say which of two theories for the same aspect of nature was the later (the more evolved) and which was the earlier . . . In this sense scientific development is a unidirectional and irreversible process, and that is not a relativistic view.[1]

This passage, however, in no way saves Kuhn from relativism. Kuhn's earlier view was seen as relativistic because that view seemed to hold that acceptance of one theory over another had nothing to do with evaluation according to neutral, objective criteria of evaluation. Proponents of competing paradigms each judged their own paradigm superior to its rivals, according to criteria their own paradigm dictated. Consequently, a given paradigm (P_1) was held to be superior according to the criteria dictated by P_1, by the proponents of P_1; while a rival paradigm (P_2) was held to be superior according to the criteria dictated by P_2, by the proponents of P_2. This rendered paradigm-evaluation relativistic in the following sense: which paradigm was judged superior depended on which set of criteria of evaluation one judged by. P_1 was superior relative to the criteria dictated by P_1, but not relative to those dictated by P_2. P_2 was superior relative to the criteria dictated by P_2, but not relative to those dictated by P_1. Since Kuhn held that there were no neutral criteria available, evaluation was relative to the criteria, hence the paradigm, appealed to. This is the sense in which his view was seen as relativistic.

The evolutionary metaphor Kuhn suggests in the above cited passage does not speak to the relativism issue just outlined. For at best, the evolutionary criteria suggested by the metaphor will indicate which of a set of competing paradigms is later, or more developed. ***But later does not mean better.*** The proponents of the earlier paradigm (as picked out according to the criteria suggested by the evolutionary construal) could still judge their paradigm superior to the later one, according to the criteria of evaluation dictated by their (earlier) paradigm. Consequently, Kuhn's evolutionary metaphor does not in the least show his view to be non-relativistic. Here, as elsewhere, Kuhn's later remarks do little to make plausible, or to clarify, refine, or improve, his original presentation.[1]

---

[1] Kuhn, 'Discussion', in Suppe [1974], p. 508.

(c) '*Objectivity, Value Judgment, and Theory Choice*'

Kuhn's most recent systematic treatment of the issues we have been considering is his 'Objectivity, Value Judgment, and Theory Choice'.[2] However, here again his arguments do little to increase the plausibility of his position.

Kuhn's presentation is largely concerned to further develop two themes already familiar to the reader: the 'no algorithm for theory choice' theme, and the 'values' theme. These we have already discussed at some length, and Kuhn adds little that requires still further inspection of this much-trodden ground. Kuhn does add a new dimension to the discussion, however. In the course of his discussion of these themes, he suggests that his critics rely on an untenable distinction—the distinction, first drawn by Hans Reichenbach,[3] between the contexts of discovery and justification. It is worthwhile, in view of the importance of this distinction, to consider Kuhn's arguments here in some detail. Kuhn doubts 'whether or not th[e] invocation of the distinction between contexts of discovery and of justification provides even a plausible and useful idealization' (p. 327) of scientific activity, and offers three reasons for this doubt.

(1) The first reason centres on what Kuhn calls the 'context of pedagogy', which is significantly distinct from the more familiar contexts we have been considering. A significant feature of science pedagogy is the tendency, Kuhn argues, to over-simplify the decisions, including decisions concerning theory choice, that scientists actually make. By relying on 'famous crucial experiments' for exemplification of scientific decisions, science textbooks and philosophy of science textbooks make scientific decisions look too easy, for crucial experiments 'are paradigms of good reason for scientific choice; they illustrate the most effective of all the sorts of argument which could be available to a scientist uncertain which of the two theories to follow' (p. 327). If crucial experiments make scientific decisions easy, Kuhn argues, they are also typically not available for scientists to consider when choosing between rival theories: 'By the time they [the crucial experiments] were performed no scientist still needed to be convinced of the validity of the theory their outcome is now used to demonstrate. Those decisions had long since been made on the basis of significantly more equivocal evidence' (p. 327). Therefore, to the (considerable) extent that philosophers of

---

[1] For related discussion, *cf.* my Siegel [1979].

[2] Kuhn [1977], pp. 320–39. The following pages have also been incorporated, with minor changes, in a paper dealing with the distinction between the contexts of discovery and of justification. *Cf.* Siegel [1980*a*]. See also the discussion of Kuhn's essay in Frankel [1978]. Frankel makes good use of the distinction between *accepting* a theory (as true) and *pursuing* a theory (as plausible, or worthy of further consideration). This distinction is in turn developed nicely in Kordig [1978].

[3] Reichenbach [1938], chapter 1. An excellent summary of Reichenbach's distinction is given in Scheffler [1967], pp. 67–73. A discussion of the current status of the distinction is presented in my Siegel [1980*a*].

science take crucial experiments to be typical reason-providers for scientific decisions, they have in mind a highly misleading picture of actual scientific decision making: 'The exemplary crucial experiments to which philosophers again and again refer . . . provides needed economy to science pedagogy, but they scarcely illuminate the character of the choices that scientists are called upon to make' (pp. 327-8).

Kuhn's historical claim, that crucial experiments do not (typically) guide scientific decisions, we will not challenge here. For even if Kuhn is correct, no ramifications for the distinction between discovery and justification follow. To see this, we need only note that scientific decisions are one thing, the justification of scientific decisions another. Scientific decision-making is, of course, a crucial sort of scientific activity; and if we are out to understand the scientific enterprise in all its aspects, we shall certainly want to study actual scientific decision-making and theory-choosing. Still, once we have a thorough account of scientist *S*'s decision to adopt theory T_1 instead of T_2, we may well want to ask the further question: was *S*'s decision justified? For what we are after, in the context of justification, is an account of the justificatory force of the scientist's reasons for adopting T_1; an account that enlightens us as to why those reasons are *good* reasons. Factors accounting for the scientist's decision to adopt T_1 need offer no such justificatory account—though they may well settle the question of why the scientist in fact made the decision he/she did. The question the context of justification is concerned to answer is not 'why did X choose T_1?', but rather 'was X's choice of T_1 a good one?' Given the legitimacy of this latter question, we must recognise that the 'context of decision' is not equivalent to the context of justification—even though, as Kuhn correctly points out, in the case of the crucial experiment the prime factor in the scientist's decision (the results of the crucial experiment) may well be the factor that justifies that decision. Thus, we can agree with Kuhn that the discovery/justification distinction does little to illuminate actual scientific decision-making, for ceding that point in no way casts doubt on the legitimacy of the distinction between the contexts.

(2) A second, related, reason for rejecting the distinction, Kuhn argues, is that typical philosophical illustrations of scientific choice distort 'by making choice [seem] totally unproblematic' (p. 328). Typically, Kuhn claims, philosophical illustrations of scientific choice situations present these situations in ways which indicate that one theory, and not the other, is clearly and obviously to be preferred: for example, in accounts of the choice between the oxygen and phlogiston theories, oxygen's advantages over phlogiston typically are presented as unequivocally clear and, seemingly, decisive; however, the advantages the phlogiston theory enjoyed over the oxygen theory are rarely taken into account. Thus, the typical portrayal of the scientist's decision to adopt the oxygen theory instead of the phlogiston theory (or, *e.g.*, the Copernican theory instead of Ptolemy's) misleads in that it fails to show that, in typical decision contexts, 'there are always at least some good reasons for each possible choice' (p. 328).

Thus far, Kuhn's point seems clear enough—decision contexts in science are rarely cut and dry. The implications of this point for the discovery/justification distinction, however, are less so. 'What', we want to ask, 'does the "non-cut-and-dryness" of decision contexts have to do with Reichenbach's distinction?' Kuhn answers as follows:

> Considerations relevant to the context of discovery are then relevant to justification as well; scientists who share the concerns and sensibilities of the individual who discovers a new theory are ipso facto likely to appear disproportionately frequently among that theory's first supporters (p. 328).

Kuhn's point seems to be this: typical theoretical decision contexts are not straight-forward or obvious; rather, each rival theoretical alternative will have reasons which support it. How scientists decide in such cases, then, depends on factors which are relevant to the discovery of the alternatives. In the oxygen/phlogiston example, factors relevant to the discovery of the oxygen theory (leaving aside for now the question of whether it makes sense to speak of theories as being 'discovered') are likely to be factors relevant to the various members of the scientific community's decisions to adopt the oxygen theory: 'scientists who share the concerns and sensibilities of the individual who discovers a new theory are *ipso facto* likely to appear disproportionately frequently among that theory's first supporters' (p. 328).

This point may be right—it may be that factors relevant to the discovery of a theory are also relevant to other scientists' decisions to adopt that theory—but it has little bearing on the justification/discovery distinction. For, whatever factors led scientists to adopt the oxygen theory (for example), we still can ask questions about the justification of that decision. As noted already, the context of justification is distinct from the 'context of decision'. So at the most Kuhn has shown that factors relevant to the context of discovery are relevant to the context of decision. He has *not* shown what must be shown in order to cast doubt on Reichenbach's distinction—namely, that factors relevant to the context of discovery are relevant to the context of justification.

(3) Kuhn offers one more reason for doubting Reichenbach's distinction. This reason depends on Kuhn's claim that 'subjective factors' are an ineliminable element of a scientist's decision to adopt a scientific theory. Much has been written concerning such 'subjective factors', and we need not enter into a detailed consideration of them here.[1] For, even if Kuhn is right that such subjective factors are an ineliminable element of a scientist's decision to adopt some theory, it nevertheless remains the case that the factors relevant to such decisions are not, necessarily, relevant to the justification of such decisions. As before, Kuhn has confused the 'context of decision' with the context of justification. And, as before, such confusion will not do. We certainly do want an account of the scientific decision-making context. But, we also want an account of other aspects of the scientific enterprise—in particular, of the justification of scientific decisions. Kuhn's reasons for denying Reichenbach's distinction all confuse decision with justification, as if deciding to adopt T is also to justify that decision. This however, is clearly a mistake—it is one thing to decide to adopt T, another to justify that decision. (If not, how could one ever make an unjustifiable decision?) Kuhn has at most shown that factors relevant to the context of discovery are relevant to the 'context of decision'; he has not shown that factors relevant to discovery are relevant to justification. Consequently, he has not provided us with good reason for rejecting Reichenbach's distinction. And so, he has not provided any reason

---

[1] For discussion of these 'subjective' factors, see my Siegel [1976], and references to Scheffler and Shapere within.

to take less seriously his critics' criticisms of his conception of scientific theory choice.

Finally, Kuhn turns his attention to the notion of 'subjectivity'. His critics err, Kuhn claims, if they take his claim that theory choice necessarily involves subjective factors to mean that theory choice is 'undiscussable' or 'a matter of taste', and those critics who take him to be claiming the contrary 'have seriously mistaken [his] position' (p. 337). Here is the 'soft' Kuhn, the Kuhn who seemingly denies the incommensurability thesis and its corollaries. Kuhn *must* be here denying the incommensurability thesis, because that thesis guarantees the 'undiscussability' of rival paradigms. According to incommensurability, rival paradigms are incapable of being meaningfully neutrally compared; proponents of such paradigms inevitably talk past each other as they speak 'across the revolutionary divide'.[1] If theory choice between such incommensurable rivals is anything, it is undiscussable. So, if Kuhn denies that theory choice is undiscussable, he must (on pain of inconsistency) deny the incommensurability thesis. Yet, one page later, Kuhn writes as if the incommensurability thesis is alive and well, as he indicates that it is 'difficult or, more likely, impossible' for 'commensurable' theory choice to take place:

> . . . I simply assert the existence of significant limits to what the proponents of different theories can communicate to one another. The same limits make it difficult or, more likely, impossible for an individual to hold both theories in mind together and compare them point by point with each other and with nature. That sort of comparison is, however, the process on which the appropriateness of any word like 'choice' depends (pp. 338–9).

We have again, thus, the by now familiar tension which runs through Kuhn's later writings: the tendency to, on the one hand, defend a strong incommensurability claim; and, on the other, deny any strong incommensurability claim. As we have seen earlier, such tension is not acceptable—Kuhn simply cannot (consistently) have it both ways. This most recent defence, therefore, does little to rectify the problems with Kuhn's view of the objectivity of theory choice already discussed.[2]

---

[1] Kuhn [1970*a*], p. 149.

[2] Others, besides Kuhn himself, have attempted to defend Kuhn from his critics. One such defence is that provided by Wolfgang Stegmüller, who, in his [1976*a*] presents a reconstruction of Kuhn's position which, Stegmüller claims, defends Kuhn from his critics, who Stegmüller labels 'the critical rationalists'. Since some philosophers have acclaimed Stegmüller's reconstruction of Kuhn (for example, Kockelmans [1976]), it is worthwhile here to point out that a positive assessment of Stegmüller's reconstruction of Kuhn is dubious, to say the least. While a comprehensive treatment of Stegmüller's treatise would take us far afield, it is possible to briefly outline the implications of Stegmüller's reconstrual of Kuhn's position for the issues presently being considered. These can be summed up as follows: Stegmüller claims that his account of scientific theories clarifies and supports Kuhn's view of the dynamics of theory change. However, concerning the issues we have been concerned with thus far—incommensurability, incomparability, relativism, rationality, and objectivity—Stegmüller's reconstrual, rather than defending Kuhn from his critics, actually grants Kuhn's critics their main points. The features of Kuhn's original presentation of theory choice, which his critics deplored —for example, the incommensurability of competing paradigms, the non-reducibility of old paradigms to their replacements, and the relativism implied by these positions—are not features of Stegmüller's reconstrual. On the contrary, according to Stegmüller's

CONCLUSION

The three recent responses by Kuhn considered above are not exhaustive, but are, I believe, representative, of Kuhn's latest attempts to clarify and refine his position concerning incommensurability and the objectivity of theory choice in science.[1] As the discussion thus far has shown, these further responses do little to render his view plausible. His position remains both confusing and confused. *The Essential Tension*, while it contains some excellent historical articles and provides a highly interesting account of the development of Kuhn's ideas, does little to settle the debate between Kuhn and his critics.

Objectivity, as Scheffler has stated, 'requires simply the possibility of intelligible debate over the comparative merits of rival paradigms'.[2] If Kuhn allows for such debate, he must indeed retract his irrationality thesis, as well as his incommensurability thesis. He claims to have done the former, while refusing to do the latter. This, it seems clear, is untenable. If he indeed retracts the irrationality thesis, he capitulates to his critics by at least granting the possibility that theory choice is objective. If he does not retract the irrationality thesis, his position is conceptually confused, and his introduction of 'values' into the discussion only further confuses the issues. His claim that there is no algorithm for theory-choice likewise offers Kuhn little help, as it both fails to clarify his view, and is at any rate agreed to by all parties in the debate. Kuhn's more recent attempts to respond to his critics do not, finally, do so adequately. The irrationality thesis, it must be concluded, remains tenuous and unwarranted; Kuhn has

---

reconstrual, normal science is not irrational (p. 158); competing paradigms are not incommensurable (pp. 37, 245); progress during theory change can be objectively measured (p. 242); old paradigms are reducible to new ones (p. 216); a philosophically objectionable relativism is not maintained (pp. 154–7); and the criteria of evaluation of competing paradigms are not paradigm-bound (p. 154). In short, virtually all the aspects of Kuhn's views that the 'critical rationalists' deplore are not aspects of Stegmüller's reconstrual of Kuhn. It seems reasonable to conclude, therefore, that Stegmüller's reconstrual does not rebut Kuhn's critics; rather, it concedes their main points, and fails to clarify or improve Kuhn's position concerning incommensurability, comparability, relativism, and the objectivity of scientific theory choice.

As indicated, this brief treatment of Stegmüller is hardly comprehensive, but deals only with Stegmüller's reconstrual as it concerns the issues being considered in this paper. The reader interested in a more detailed discussion of Stegmüller's overall view is directed to his [1976*a*], and to the review of that book by Joseph J. Kockelmans (Kockelmans [1976]). Kuhn's attitude towards Stegmüller's reconstrual is given in his [1976]. Stegmüller's more recent developments of his view are presented in his [1975] and his [1976*b*].

[1] Besides the responses considered thus far, it is appropriate to mention two others here. Kuhn's 'Second Thoughts On Paradigms', originally published in Suppe [1974] and reprinted in Kuhn [1977], pp. 293–319, attempts to refine his conception of paradigm. While space forbids detailed consideration of the point here, I believe it can be shown that Kuhn's refinement of the paradigm concept in that paper substantially reduces the support that concept originally gave to the incommensurability thesis, thus further weakening the already questionable status of the incommensurability thesis discussed above. I have briefly argued for this point in my Siegel [1980*b*]. In addition, Jack W. Meiland, in his [1974], attempts to construe Kuhn as a proponent of objectivity in science. I have not included discussion of Meiland's claims in the present paper because I have elsewhere replied in detail to Meiland's analysis, *cf.* my Siegel [1976].

[2] Scheffler [1972], p. 369.

failed to show that theory choice cannot be objective or rational. On the other hand, if Kuhn seriously agrees with his critics that science is not necessarily subjective or irrational, as he has claimed, he has failed to show how this sentiment can be maintained in conjunction with the rest of his views concerning the dynamics of theory change.[1]

HARVEY SIEGEL
University of Nebraska, Lincoln

REFERENCES

FRANKEL, H. [1978]: 'Review of T. S. Kuhn, *The Essential Tension*', *Philosophy of Science*, **45**, pp. 649–52.
KITCHER, P. [1978]: 'Theories, Theorists, and Theoretical Change', *Philosophical Review*, **87**, pp. 519–47.
KOCKELMANS, J. J. [1976]: 'Review of W. Stegmüller, *The Structure and Dynamics of Theories*', *Philosophy of Science*, **43**, pp. 293–7.
KORDIG, C. R. [1971]: *The Justification of Scientific Change*. Dordrecht, Holland: D. Reidel Publishing Company.
KORDIG, C. R. [1978]: 'Discovery and Justification'. *Philosophy of Science*, **45**, pp. 110–17.
KUHN, T. S. [1957]: *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*. Cambridge, Mass.: Harvard University Press.
KUHN, T. S. [1969]: 'Postscript – 1969'. Included in Kuhn [1970*a*].
KUHN, T. S. [1970*a*]: *The Structure of Scientific Revolutions*. First edition, 1962; second edition, enlarged. Chicago: University of Chicago Press.
KUHN, T. S. [1970*b*]: 'Reflections on my Critics'. Included in Lakatos and Musgrave [1970], pp. 231–78.
KUHN, T. S. [1971]: 'Notes on Lakatos', in R. Buck and R. Cohen (*eds*.): *Boston Studies in the Philosophy of Science*, vol. 8, Dordrecht, Holland: D. Reidel Publishing Company.
KUHN, T. S. [1976]: 'Theory-Change as Structure-Change: Comments on the Sneed Formalism', *Erkenntnis*, **10**, pp. 179–99.
KUHN, T. S. [1977]: *The Essential Tension*. Chicago: University of Chicago Press.
LAKATOS, I. and MUSGRAVE, A. (*eds*.) [1970]: *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.
LAUDAN, L. [1976]: 'Two Dogmas of Methodology', *Philosophy of Science*, **43**, pp. 585–97.
LEVIN, M. E. [1979]: 'On Theory-Change and Meaning-Change', *Philosophy of Science*, **46**, pp. 407–24.
MEILAND, J. W. [1974]: 'Kuhn, Scheffler, and Objectivity in Science', *Philosophy of Science*, **41**, pp. 179–87.
MOBERG, D. W. [1979]: 'Are There Rival, Incommensurable Theories?', *Philosophy of Science*, **46**, pp. 244–62.
QUINE, W. V. [1960]: *Word and Object*. Cambridge, Mass.: M.I.T. Press.
REICHENBACH, H. [1938]: *Experience and Prediction*. Chicago: University of Chicago Press.
SCHEFFLER, I. [1967]: *Science and Subjectivity*. New York: Bobbs-Merrill.
SCHEFFLER, I. [1972]: 'Vision and Revolution: A Postscript on Kuhn', *Philosophy of Science*, **39**, pp. 366–74.
SHAPERE, D. [1964]: 'The Structure of Scientific Revolutions', *Philosophical Review*, **73**, pp. 383–94.
SHAPERE, D. [1966]: 'Meaning and Scientific Change', in R. G. Colodny (*ed*.): *Mind and Cosmos*, Pittsburgh: University of Pittsburgh Press, pp. 41–85.
SHAPERE, D. [1971]: 'The Paradigm Concept', *Science*, **172**, pp. 706–9.
SIEGEL, H. [1976]: 'Meiland on Scheffler, Kuhn, and Objectivity in Science', *Philosophy of Science*, **43**, pp. 441–8.

---

[1] I should like to thank Israel Scheffler for helpful comments on an early draft of this paper.

SIEGEL, H. [1980]: 'Epistemological Relativism in its Latest Form', *Inquiry*, **23**, pp. 107–117.
SIEGEL, H. [1980*a*]: 'Justification, Discovery, and the Naturalizing of Epistemology', *Philosophy of Science* (forthcoming).
SIEGEL, H. [1980*b*]: 'Rationality, Talking Dogs, and Forms of Life', *Educational Theory* (forthcoming).
STEGMÜLLER, W. [1975]: 'Structures and Dynamics of Theories: Some Reflections on J. D. Sneed and T. S. Kuhn', *Erkenntnis*, **9**, pp. 75+100.
STEGMÜLLER, W. [1976*a*]: *The Structure and Dynamics of Theories*, translated from the German *Theorienstruckturen und Theoriendynamik* by Dr William Wohlheuter. New York: Springer-Verlag.
STEGMÜLLER, W. [1976*b*]: 'Accidental ("Non-Substantial") Theory Change and Theory Dislodgement: To What Extent Logic Can Contribute to a Better Understanding of Certain Phenomena in the Dynamics of Theories', *Erkenntnis*, **10**, pp. 147–78.
SUPPE, F. (*ed.*) [1974]: *The Structure of Scientific Theories*. Urbana: University of Illinois Press.

REICHENBACH'S THEORY OF REASONABLE ASSERTION*

1 *Introduction*
2 *Laws and Modalities*
3 *The Theory of Admissible Statements*
4 *Counterfactual and Predictive Conditionals*
5 *Conclusion*

1 INTRODUCTION

Our willingness to assert sentences of natural language often fails to correspond to our knowing precisely those facts that are prescribed as the truth-conditions of those sentences by the standard interpretation of logical connectives. There arises then the question of how the conditions of reasonable assertion may be defined. Noting that the problem is not confined to the use of the subjunctive mood or the conditional form of sentence, Hans Reichenbach in his [1947] attempted to furnish a general solution. His posthumously published [1954] significantly revises and expands the earlier treatment and is now available as *Laws, Modalities, and Counterfactuals*, a title which reflects that the work includes an explication of the notion of scientific law. (My parenthetical page citations, when not otherwise identified, will refer to this book.)

There appear to be essentially two ways of viewing the problem of reasonable assertion. According to what may be called the *variant-meaning approach* the problem arises because at least some logical connectives are mainly used with meanings that differ from those of their counterparts in standard formal logic. The variant-meaning theorist may attempt to explicate these meanings, and an important question for him is whether the explication can be furnished within a truth-functional language. The alternative viewpoint might be referred to as the *variant-assertibility approach.* On this view one takes the meanings of the connectives of ordinary language to be essentially those prescribed by the standard logical interpretation, and one tries to show how the reasonable assertibility of a sentence so construed may depend on factors other than its truth. The variant-assertibility theorist may, for example, point out that there is a recognisable sense of 'reasonable' according to which it is unreasonable for a person to assert a contingent sentence S when he is in a position to assert an obviously simpler sentence S' which implies S. The problem on this view is to furnish a fuller and more precise account along such lines.

---

* Review of Hans Reichenbach [1977]: *Laws, Modalities, and Counterfactuals*. Berkeley: University of California Press. £8.05. Pp. 140.

Reichenbach's analysis does not wholly fit into either of these approaches but seems rather to involve a mixture of the two. On the one hand he holds that the connectives of natural language are ambiguous, sometimes being used as in standard logic and sometimes being used with other meanings which he proposes to explicate. The explication is to be accomplished by stating in a truth-functional metalanguage the truth-conditions for sentences whose connectives are meant in the non-standard way. (He refers to connectives used with such meanings as 'reasonable operations'.) This seems to make him a variant-meaning theorist. On the other hand he indicates that there are degrees of reasonableness, and he holds that some law statements known to be true become reasonable only within certain wider knowledge situations, and his theory of counterfactuals must be viewed as a theory defining the conditions under which true conditionals may reasonably be asserted counterfactually in given knowledge situations. These and other things suggest the variant-assertibility approach. He seems, however, to have thought that technical means for distinguishing intuitively reasonable from intuitively unreasonable assertions could be provided to a sufficient degree by a theory concerned with imposing conditions on sentences. That is, a theory of reasonable assertion may be obtained through a theory of reasonable operations (connectives). How this works out should become clearer in later sections of this paper.

2 LAWS AND MODALITIES

The first step Reichenbach takes toward a theory of reasonable operations is to formulate a theory of nomological statements. (I shall follow Reichenbach in using 'sentence' and 'statement' interchangeably.) The relatively precise term 'nomological statement' is introduced as an explication of the ordinary notion 'law of nature or of logic'. The explication is not intended to throw any additional light on the notion of analyticity, however, and is therefore essentially a *theory* of synthetic nomological statements (laws of nature). The theory is supposed to involve no modal concepts such as physical necessity or possibility and no use of non-truth-functional connectives. In particular it is important to note that logical connectives occurring in nomological statements are, within the context of the theory, to be interpreted in the standard way.

Reichenbach first defines a class of statements that he calls *original* nomological statements; a nomological statement as such is any logical consequence of this class. The essential strategy for defining original nomological statements can, I believe, be viewed as consisting of two steps. First, certain requirements are laid down which suffice to insure that any synthetic statement meeting them is not capable of being *conclusively* verified. Therefore, any statement meeting these requirements, if capable of being confirmed to an appreciable degree, will be subject to confirmation only by methods that are, broadly speaking, inductive. Second, it is required that an original nomological statement be 'verifiably true', which means it must be not only true but also actually confirmed to a high degree at some time or other. The only concept that the explication requires outside the domains of logic and semantics is the concept of confirmation, an analysis of this having been furnished to Reichenbach's satisfaction in his [1949].

Even if one grants that the notion of confirmation is less mysterious than that of law one might question the explication on the grounds that there may be laws of nature that are never confirmed. It would seem that the notion required is not that of being confirmed but rather that of being confirmable. Reichenbach recognises the objection and devotes a separate chapter to the task of explicating in non-modal terms the notion of being 'verifiably true in the wider sense', *i.e.*, true and confirmable. I have argued elsewhere [1977] that this attempt is not successful. It may then be that Reichenbach's achievement is at best that of having successfully demarcated the class of scientific laws that become known at some time or other. While this issue is important for the problem of explicating law I think it is much less so for the theory of reasonable assertion. Insofar as the problem of law bears on the problem of reasonableness what is relevant is surely some set of known laws or even some set of statements merely taken to be laws. Reichenbach himself mentions the notion of being reasonable relative to a given system of assumed laws (p. 84). Near the end of the book (p. 120) he makes it clear that the set of reasonable statements that is his primary concern is the set based on the system of laws that are actually confirmed at some time or other. It appears then that he did not regard his wider concept of verifiability as vital for his theory of reasonable operations.

Since any statement following from a nomological statement is nomological, laws of nature on this explication will include many non-general sentences such as 'If this rod is made of copper then this rod is electrically conductive'. This feature may be viewed as an undesirable departure from our usual idea of a law of nature as general and free from individual reference. Perhaps then only a suitable subclass of nomological statements can properly be regarded as laws of nature. Even so, it is nomological statements in general that appear to be required in at least two important applications of the theory: the explication of the physical and logical modalities, and the theory of reasonable operations.

Reichenbach's definitions of the modalities are taken over intact from his [1947] and run rather as one would expect. Thus, the fact that it is ***physically necessary*** that if this rod is made of copper then it is electrically conductive is explicated by the metalinguistic sentence 'The statement "If this rod is made of copper then it is electrically conductive" is a synthetic nomological statement'. And so forth. But Reichenbach commendably recognises that most of our modal locutions are not of this 'absolute' type but rather are 'relative'. For example, knowing that a particular rod is made of copper, one might just assert that it is physically necessary that the rod is electrically conductive. The last clause here is not itself nomological but follows from the conjunction of a nomological statement and a particular fact. It is 'nomological relative to' that fact.

The first application Reichenbach makes of his theory of nomological statements is in his definition of 'three orders of truth', (p. 66), which could perhaps have more fittingly been called 'three modes of access to truth'. Logical truths and analytic sentences are said to be true of third order. Synthetic nomological statements are true of second order. Verifiably true sentences that are not nomological are true of first order. These definitions play an important role in the theory of admissible statements, a topic to which I shall now turn.

3 THE THEORY OF ADMISSIBLE STATEMENTS

Reichenbach sometimes writes (pp. 5 and 67) as though he considered the explication of reasonable operations to be furnished wholly by his theory of admissible statements. That this is not so is later (p. 101) recognised in a revised statement. The theory is nevertheless central to his work, and it forms that part of his theory of reasonable assertion that is roughly in accord with the variant-meaning approach.

Admissible statements form a certain subclass of nomological statements and, like the latter, their logical connectives are to be interpreted in the standard way. Admissibility is, however, to be a sufficient condition for reasonableness in the following way: if a given statement is admissible then its major connective can also be interpreted as a 'reasonable operation', *i.e.*, the statement is also true when its major connective is interpreted in the 'reasonable sense'. (Connectives other than the major connective of a sentence are generally to be construed in the standard way.) The requirements for admissible status are untidily complex; Reichenbach's presentation of the ideas involved is scattered over a number of pages (pp. 19–30, 48, 66–7, and 70–1). I shall here attempt mainly to convey some idea of what they do and do not accomplish.

One way in which the requirements for admissibility contribute to intuitive reasonableness is in ruling out many unnecessarily complicated sentences in favour of logically equivalent sentences of a simpler kind. For example, the sentence 'If all metals are heavier than sulfur then all metals expand when heated, and it is also the case that either all metals are heavier than sulfur or else all metals expand when heated' is not admissible, but its logical equivalent 'All metals expand when heated' is. In this way many sentences suggesting the existence of connections that are in fact irrelevant are ruled out. Notable exceptions exist. For example, the admissible statement 'If that rod was metal and was heated then it expanded' has as an equivalent variant that is also admissible the sentence 'If that rod was metal it expanded or if that rod was heated it expanded'. This odd result may partly be accounted for by recalling that admissible status is supposed to insure the reasonable sense of only the major sentence connective, which in this case is 'or'.

The requirements also contribute to intuitive reasonableness by ruling out (in many cases) a nomological statement true of a given order in favour of a simpler but stronger statement that is true of the same order. Thus, 'All copper rods that are electrically nonconductive are purple', which is true of second order, is ruled out in favour of 'All copper rods are electrically conductive', which logically implies it. Similarly, the third-order sentence 'Everything that is both round and square is hot' is ruled out in favour of 'Nothing is both round and square'. That this is not simply a matter of ruling out necessarily vacuous antecedents or the like can be seen by considering the intuitively unreasonable but true sentence 'For every signal *x* there is a signal *y* such that if *x* is faster than *y* then *x* has a higher frequency than *y*' in which both the antecedent and consequent are instanced by some but not all ordered pairs of signals. It is not admissible, being ruled out in favour of 'For every signal *x* there is a signal *y* such that *x* is not faster than *y*', which implies it. This last follows from the law that no signal is faster than light, but it is not ruled out in favour of the latter, both being admissible statements.

The theory of admissible statements is largely successful in selecting from the class of quantified nomological statements those that are intuitively reasonable in the sense that one possessed of the requisite general scientific knowledge would normally be willing to assert them. The theory is much less successful with the class of molecular nomological statements, but I should like to pass over this in order to dwell on a problem that is even more important. It is fairly obvious that the vast majority of the sentences actually asserted are not nomological. Suppose that, speaking of a certain jewel, someone asserts 'Alice stole it or Betty stole it' (or equivalently 'If Alice didn't steal it then Betty did'). The person's willingness to assert this might well stem from his knowing that somebody did steal the jewel and that nobody except Alice and Betty had the opportunity. Now the sentence 'If somebody stole the jewel and only Alice and Betty had the opportunity to steal it, then either Alice stole it or Betty stole it' is indeed nomological and also admissible. But the sentence actually asserted is neither. In Reichenbach's terminology the asserted sentence is 'admissible relative to' the background information mentioned, but relative admissibility of this kind is by no means sufficient for reasonableness. If one's background information consisted simply of the fact that Alice stole the jewel, then the assertion of 'Alice stole it or Betty stole it' (or of 'If Alice didn't steal it then Betty did'), although true, would not be reasonable. And this is so in spite of the fact that 'If Alice stole the jewel then Alice stole it or Betty stole it' is an admissible statement. The theory of admissible statements furnishes no means for deciding the reasonableness of many similar assertions that play a prominent role in our everyday discourse.

Another matter to which Reichenbach might well have devoted more attention is the conditions under which *denials* are reasonable. The denial 'It is not true that if this rod is copper then it is soluble in water' seems reasonable even when its logical equivalent 'This rod is copper and is not soluble in water' is not (*e.g.*, when the rod is made of iron). In his [1947] Reichenbach suggests (p. 380) that in everyday usage the denial of a conditional is often to be construed as equivalent to a metalinguistic sentence denying that the conditional in question is a nomological statement. This is doubtless a step in the right direction, but it is hardly plausible (especially in view of the considerations of the last paragraph above) to suppose that what is usually being denied is precisely the *nomological* status of the statement concerned. It seems safe to say that in everyday usage the ostensible denial of a conditional, biconditional, or disjunction is frequently to be construed rather as a denial of the reasonableness of asserting the sentence in question.

In constructing the theory of admissible statements Reichenbach seems to have had primarily in mind the application of the theory to the problem of counterfactual and predictive conditionals. He may have supposed that any theory of reasonable assertion capable of handling that recalcitrant problem should suffice for the resolution of other less celebrated aspects of the problem of reasonableness. That does not seem to be the case. I shall now turn to the major application of the theory of admissible statements, the theory of counterfactual and predictive conditionals.

4 COUNTERFACTUAL AND PREDICTIVE CONDITIONALS

Within the class of counterfactual conditionals Reichenbach distinguishes between those he calls 'regular' and those he calls 'counterfactuals of noninterference'. Postponing a discussion of the latter, I shall refer to the former kind simply as counterfactuals. Counterfactuals are sentences such as 'If Bill had drunk the whole bottle he would have become ill' and 'If Harry were here Betty would be happy'. The grammatical form of the counterfactual signals that the asserter believes the antecedent of the corresponding material conditional to be false. Like most writers on the subject, Reichenbach assumes that the falsity of the consequent is also indicated. As Parry ([1957], p. 93) has pointed out, this is not so. But in any case the asserter will certainly not believe that the consequent is true, and the assumption that disbelief in both antecedent and consequent is indicated appears to be harmless for the purpose of devising a theory. Counterfactuals of the kind of concern here should, I believe, be distinguished from sentences of a less problematic kind that might be called 'purely inferential counterfactuals'. I am thinking of sentences such as 'If Harry were here now he would have to have averaged over 80 miles an hour' and 'If that match had not lit it could not have been dry'. A distinction of this kind has been noted by Parry [1957] and Sellars [1958], who have also observed that the more problematic type frequently has causal connotations that are lacking in the kind I am calling purely inferential. Reichenbach does not make any such distinction, but his examples show that he is at least primarily concerned with the more problematic type.

Sentences such as 'If you were to touch this wire you would receive a shock' are not counterfactuals, this being equivalent to 'If you touch this wire you will receive a shock'. It is rather an example of a *predictive* conditional, most predictive conditionals being indicative conditionals whose consequents are in the future tense. Aside from appropriate assumptions concerning the truth-values of antecedent and consequent Reichenbach apparently intends his analysis of counterfactuals to furnish also the conditions of reasonableness for such conditionals. I shall therefore concentrate on his theory of counterfactuals.

The problem of counterfactuals as Reichenbach sees it can perhaps be identified with the question of what conditions imposed on the corresponding material conditional are necessary and sufficient for the reasonable assertion of a counterfactual. He holds that a sufficient condition is that the material conditional be an admissible statement. Otherwise it is necessary and sufficient that it be relative-admissible in a certain way. The main problem is viewed as that of stating the further conditions which together with relative-admissibility are just those required for reasonable assertibility.

If a conditional $B \supset C$ is relative-admissible then by definition there is a sentence A such that $A \supset (B \supset C)$ is an admissible statement. Reichenbach holds that whether a counterfactual corresponding to $B \supset C$ can reasonably be asserted depends on whether certain conditions involving A and B are satisfied. Before examining these, let us look at some particular cases in which the relative admissibility requirement is satisfied. In each of the following situations the first condition mentioned plays the role of A above and the counterfactual asserted corresponds to the material conditional $B \supset C$. The cases to be considered are the following: (1) One knows that a rod is made of metal (and was not heated, did not expand), and one asserts 'If that rod had been heated it would have expanded'; (2) One knows that a rod was heated (was not made of metal, did not expand), and one asserts 'If that rod had been made of metal it would have expanded'; (3) One knows that a rod did not expand (was not heated, was made of metal), and one asserts 'If that rod had been heated it would not have been made of metal'; (4) One knows that a rod did not expand (was not made of metal, was heated), and one asserts 'If that rod had been made of metal it would not have been heated'. Of these, (1) and (2) seem clearly reasonable while (3) and (4) seem clearly unreasonable. (It is worth noting that in each of these cases if the parenthetical information is deleted then the corresponding past-tense indicative conditional *is* reasonable.)

Given, then, an admissible statement of the form $A \supset (B \supset C)$ for which one knows B and C to be false, what further conditions are necessary and sufficient for the reasonable assertion of the counterfactual corresponding to $B \supset C$? Reichenbach's answer is that there are two: (I) The probability of A must be high (very close to 1), and (II) The probability of A with respect to B must be at least as great as the probability of A. These requirements must, however, be understood in the context of Reichenbach's frequency theory of probability according to which the term 'probability' is here being used in a fictitious sense, probability being, properly speaking, a property not of individual events or propositions but rather of classes or sequences of such entities. Thus, a claim that it is highly probable that a particular rod did not expand at a given time is to be interpreted as the claim that within a certain reference class of situations the relative frequency with which rods expand is quite low. And this does not mean merely that there *exists* such a class of situations; rather, the statement is regarded as making implicit reference to a particular class. Reichenbach elsewhere ([1949], pp. 374 ff.) provides general advice on how to choose reference classes so that one's probability statements may be fit to play the sort of role normally expected of them. In any given application, then, requirements (I) and (II) are to be regarded as elliptical statements implicitly referring to a particular reference class, one that has been judiciously chosen in accordance with Reichenbachian advice as that applies to the particular knowledge situation of the person making the assertion. I take it as evident that most counterfactuals actually asserted do not make even implicit reference to persons or knowledge states, and so Reichenbach could not consistently have regarded his requirements as furnishing *truth-conditions* for counterfactuals, although it is not clear that he realised this. The requirements can, however, be viewed as conditions for the reasonable assertion of counterfactuals, and this is the viewpoint I have already taken above.

Reichenbach makes it clear that in any given application it is one and the same reference class that is implicitly referred to in (I) and (II), and he refers to this class as G. He indicates that G is to be modelled on the actual situation in which the event or condition described by the statement *A* arose, although the description used for *G* is not to include *A* itself. It may, however, include 'all kinds of evidence' we have for the statement *A* (pp. 96–7). And I think we can glean from later developments (pp. 103 ff.) that *G* is to be logically compatible with both the truth and falsity of the statements *A*, *B* and *C*.

Let us apply Reichenbach's theory to the hypothetical cases (1)–(4) above. Requirement (I) is clearly satisfied in all of these; the question is whether requirement (II) enables us to pick out the intuitively acceptable cases. It seems clear enough that cases (1) and (2) satisfy (II), but whether (3) and (4) do is not so obvious. Keeping the frequency interpretation in mind, the question in (3), for example, would seem to involve a comparison between the relative frequency with which rods in 'similar' situations fail to expand and the relative frequency with which *heated* rods in such situations fail to expand. It seems plausible that heated rods fail to expand less frequently in similar situations than do rods in general. If so, (II) is not satisfied and (3) is to be rejected. (Note that we are not to make use of the information that the rod was metal—see the last paragraph above.) Similarly, it appears that (4) fails to satisfy (II) and must also be rejected. The theory thus appears to furnish results that are in accord with our intuitions.

This favourable verdict must, however, be questioned in view of Reichenbach's insistence that not only is the same class *G* referred to in (I) and (II) but *G* is also to be modelled on the actual situation to the point of including the evidence one has for the statement *A*. Referring to (3) again, the aspect of the actual situation conferring high probability on the non-expansion of the rod might well be that the rod was undergoing careful observation or precise measurement at the time in question and that no indication of expansion was produced. But then the class of 'similar' situations would consist of situations in which such observation or measurement produced no such indication. Now, it is plausible that the relative frequency with which rods expand in *that* class of situations is approximately zero regardless of whether they are being heated. But now the probability relations needed for the rejection of (3) and (4) do not seem to be forthcoming. In general, if requirement (I) is to be satisfied then one cannot exclude such observational evidence from the reference class, since such evidence will frequently be the only kind one has. It appears, then, that the only option is to drop the insistence on the identity of the implicit reference in (I) and (II). For (3), then, the implicit reference class for (II) might be the class of situations in which a rod is under observation—in some of these the rod will be heated, in some it will be made of metal, in some it will expand, *etc.* Within such a class we seem to have the desired circumstance that the relative frequency with which heated rods fail to expand is less than that for rods in general. As applied to (3), then, requirement (II) is intuitively the requirement that the probability of a rod's not expanding, given that it is heated, is at least as great as the probability that a rod in general is not expanding. Since this is not true, the counterfactual in (3) is to be rejected.

Aside from whatever qualms one may have about using the notion of probability for solving the problem of counterfactuals, Reichenbach's (I) and (II), understood in the way just suggested, appear to furnish intuitively correct rulings in an impressive variety of cases. But imagine a kind of slot machine which upon receiving the deposit of a suitable coin grants one the privilege of pulling its lever three times, each such pull resulting in a display of a colour: red, or yellow, or blue. Suppose the construction of the machine to be such that the probability for the display of any given colour on any given pull is 1/3, with the following exceptions: (*a*) If red appears on the first pull then the probabilities for the second pull are 2/5 for red, 1/5 for yellow, and 2/5 for blue, and (*b*) A display of red on the first pull and yellow on the second is invariably followed by blue on the third. This last may be viewed as a result of the fact that for some suitable description *D* of the machine's construction the sentence 'If *D* and pull #1 yields red, then if pull #2 yields yellow then pull #3 yields blue' is an admissible statement. Suppose now that one has deposited one's coin and has pulled the lever three times, resulting in the colour sequence red–blue–red. It would now be reasonable to assert 'If yellow had been displayed on my second pull then blue would have been displayed on my third'. Checking this by Reichenbach's criteria, (I) is clearly satisfied if we observed that red was in fact displayed on the first pull. If requirement (II) is understood in the way I have suggested it must be, it will be satisfied here just in case the probability of a red display on a first pull, given that yellow is displayed on the next pull, is at least as great as the probability of a red display on the first. The latter is of course 1/3, and the former can be calculated with the aid of Bayes's theorem to be 1/15. Requirement (II) is therefore not met, a result that conflicts with the obvious reasonableness of the assertion in question. One might, of course, reject the interpretation of (II) used here, but then it is difficult to see how the assertions in (3) and (4) and many other unreasonable counterfactual assertions are to be ruled out. I believe this example shows that the probability relation of (II) is not actually involved in determining the reasonableness of counterfactuals. It may, however, in some way reflect some of the actual determinants of reasonableness so as to make it symptomatically useful.

What Reichenbach calls a 'counterfactual of noninterference' is produced when, knowing that *B* is false and *C* is true, one asserts 'If *B* had been true *C* would still be true'. Such sentences were called 'semifactuals' by Goodman [1947], who considered them to be equivalent to the denials of the corresponding counterfactuals. As Parry ([1957], p. 93 f.) has pointed out, this view is not tenable. There could therefore be a distinct problem concerning these sentences. Reichenbach analyses these without the aid of the theory of admissible statements, holding that 'If *B* had been true *C* would still be true' is reasonable just in case the probability of *C* with respect to *B* is at least as great as the probability of *C*. I think this is not a sufficient condition, as can perhaps be seen from the example of Jones, who bought several tickets in a lottery and did not win. The probability of his losing, given that he bought only one ticket, is at least as great as the probability of his losing in general, but one would not therefore assert 'If Jones had bought only one ticket he would still have lost'. Actually, there seems to be no good reason for supposing that the basic determinants of the reasonableness of counterfactuals of this kind are significantly different from those relevant to the reasonableness of regular counterfactuals.

5 CONCLUSION

The basic weakness of Reichenbach's work on reasonable assertion is, I think, the failure to distinguish between what I have called the variant-meaning and the

Brit. J. Phil. Sci. 31 (1980), 385–416 *Printed in Great Britain*

Reviews

FEYERABEND, PAUL [1978]: *Science in a Free Society*. London: New Left Books. £7.50. Pp. 221.

In this collection of essays Feyerabend extends the criticism of theories of scientific rationality begun in *Against Method* and pays increasing attention to the problem of the proper role of scientific experts in a democratic society. As usual Feyerabend adopts the role of peripatetic picaroon—Professor Gellner is chastised for his 'tierischer Ernst' while Genghis Khan is praised for his advanced views on religious tolerance. And we are treated to *PKF*'s horoscope as well as the story of his intellectual development.

But did I detect a slight mellowing in the Til Eulenspiegel of contemporary philosophy? Feyerabend now calls for a '*slow* erosion of the authority of science' [his italics] (p. 102) and describes the kind of society which he favours as 'an assembly of mature people' (p. 87) and emphasises that mature decision making must be learned and requires hard work (p. 97). Nevertheless, people wearing a pacemaker should approach the book with caution and may wish to close their eyes when they come to the horrible bits. For me the scariest sentence was the following: ' "Objectively" there is not much to choose between anti-semitism and humanitarianism' (p. 9). Luckily, this episode occurred early on and after that I, at least, was able to relax. Some of my friends and colleagues who are called 'uninformed intellectual fascists', 'illiterates', or 'janitors in the Popperian temple' may find it tougher going.

Perhaps the last section of the book, a series of rejoinders to reviewers of *Against Method*, should be read first. (These replies have already appeared in various journals.) Here Feyerabend not only clarifies the content of his earlier arguments but also explains and defends his way of arguing. (See especially his 'Guide for the Perplexed'.) He holds that philosophical pronouncements, rather like recommendations to the lovelorn, are to be evaluated within a specific context. ('Look before you leap' would be good advice for Hera, but inappropriate for J. Alfred Prufrock.) Thus, a Feyerabendian position such as epistemological anarchism is to be viewed primarily as a specific remedy for today's philosophy of science, not necessarily as an elixir for all seasons. For this reason, in evaluating Feyerabend's assertions I will always try to relate them to a particular historical context.

1 REASON AND PRACTICE

Everyone agrees that the content of scientific theories changes over time. Many philosophers would also claim that science progresses because as measured against certain invariant, objective *standards*, later theories are (in general) better than earlier ones. Furthermore, philosophers believe that the practice of science is *rational* (for the most part) because scientists employ methods which help them improve their theories.

Feyerabend has mounted a series of attacks on the thesis that science is (or ought to be) governed by immutable standards and methods. In *Against Method* he argued that since great episodes in science frequently involve hypotheses which are *ad hoc*, content decreasing, or inconsistent with both data and other well established theories, current philosophies of science are not only hopelessly out of touch with real live scientific practice, but would actually kill science if taken seriously.

The obvious parry is to say that philosophical theories of science are to be understood as including a *ceteris paribus* clause. ('Don't make *ad hoc* modifications of your system—*unless* you can't think of anything else to do and please try to improve them as soon as possible.') This sort of line is adopted by Worrall in a review of *Against Method.*

Feyerabend ripostes that this is 'quite a concession for Popperians to make and it took them a long time to make it' (p. 213). As a matter of fact such admissions are scattered throughout Popper's essays. (For example, in *Conjectures and Refutations* he describes *ad hoc* hypotheses as 'stop-gaps' (p. 287) and refuted theories as 'stepping stones' or 'instruments' (p. 245). Obviously Popper doesn't like such theories, but he would not excommunicate them.)

Feyerabend also claims that if we weaken methodological rules with *ceteris paribus* clauses and talk of what one should do in the long run, they become practically empty. The advice which they offer to the scientist *at the time* at which methodological decisions must be made comes very close to 'anything goes'. I agree that such rules are rather weak, but would maintain that they are nevertheless useful. When climbing mountains, it is a good rule to choose the path going up, other things being equal. (Of course, we may occasionally have to traverse down a rock face, *reculer pour mieux sauter.*)

Much of this controversy can be undercut by clearly distinguishing (as Worrall does) between *appraisal* and *advice.* For example, severe experimental tests are always cognitively preferable to routine repetitions of old experiments. However, there may be good pragmatic, moral, or even theoretical reasons for not carrying out a severe test in certain circumstances. Thus the practical advice one would give is very dependent on the historical context while the philosophical appraisal of tests is not.

In the new essays in this volume Feyerabend develops a different attack on methodology. There can be no immutable set of standards for science because standards depend on cosmologies (and it is granted by all concerned that world views change).

Feyerabend is clearly right in certain cases. The current requirement of covariance in physics will probably stand or fall with relativity theory. It could even be argued that both the method of analysis (roughly, divide both problems and physical systems into parts and tackle them sequentially) and the method of ideal laws depend on cosmological assumptions about independence and the separability of essential and accidental properties. But before reading Feyerabend, I would have maintained that desiderata such as high empirical content and predictive power were universal (or at least components of every cosmology compatible with science).

Now I am not so sure. Feyerabend argues that within the Ptolemaic astronomical programme of 'saving the phenomena' it is not desirable that new theories make novel predictions. And given the Aristotelian view of a world which is finite both qualitatively and quantitatively, it is silly always to demand deeper explanations with higher content. It is only when people adopt a cosmological belief in an infinite universe with as yet undiscovered 'Americas of Knowledge' and an invisible microstructure that the methodological requirements of depth, predictive power, and increasing content make any sense (pp. 60–1). Feyerabend could maintain that, at best, contemporary philosophers are mapping out the contours of our present world view (or more likely, given their general ignorance of technical science, a slightly outdated one). But since great advances in science require drastic metaphysical changes, such philosophers are like generals who figure out better and better methods of fighting the last war.

My response is to grant that methodology is more metaphysics-dependent than philosophers generally like to admit—Popper's principle of methodological individualism is a case in point. Nevertheless, there are several defences of the current philosophical enterprise which can be made.

First of all, surely some rules and standards are rather like logic—they are appropriate for all possible worlds. Perhaps the best candidate for a universal standard is that of consistency. Feyerabend argues, however, that not only is this requirement constantly breeched in practice, but also that one favourite argument for it, namely that from a contradiction everything follows, assumes that one is using traditional formal logic, which physicists obviously don't! My impression is that although scientists frequently work with theories which they have good reason to believe are literally false (this is why we need a good philosophical account of verisimilitude), it is quite rare for scientists to tolerate systems which are internally inconsistent and when they do, resolving the inconsistency is a top priority problem.[1]

Secondly, it could be argued that certain methodological principles can at least be vindicated in Feigl and Reichenbach's sense; *i.e.*, they may not always work, but they'll work if anything will. Thus a proponent of determinism (or causality, or realism, or high content), could argue as follows: The structure of the world may not be deterministic (or causal, *etc.*). However, we are well-advised to *look* for universal laws—if we settle for something weaker, we'll never find them if they do exist. Of course, such a position does not justify a ban on non-deterministic hypotheses.

Thirdly, the methodologist could adopt a stance similar to that of Carnap who, after discovering a continuum of inductive methods, proposed that experience be used to choose the c-function matched to this world. Thus, it could be argued, there is no *a priori* way to choose between Aristotle's desiderata for science and those of modern methodology. It is only the *post hoc* success of modern science which proves that our methodology is better. Exactly, Feyerabend might reply—methodological rules are formed through scientific practice; they do not precede it. However, even if we admit that methodological rules arise dialectically, we could still defend the current philosophical interest in formulating them on two grounds. First, the aspects of cosmology which influence methodology change relatively slowly. Basic methodological norms are rather like Pópper's 'basic' statements—they are not immune to criticism but in most scientific disputes or decisions they can be taken for granted. Secondly, I have suggested elsewhere[1] that even when standards do change, there is typically a sort of correspondence relation between the old rules and the new; *i.e.*, we can both show how the new rules are better and also explain why the old rules led to limited progress.

---

[1] See the historical examples in my [1970].

To sum up, Feyerabend convinced me that most methodological rules are based on metaphysics, not pure logic. Therefore, they, like metaphysics, can be revised as science grows. Furthermore, even in an epoch where the basic world view is not being challenged, these methodological rules should never be taken as rigid absolutes. At most they are like the *prima facie* obligations of moral philosophy to keep our promises, *etc.* It is easy (thanks to Feyerabend's historical case studies) to produce examples where it was reasonable to relax them. He did not convince me that such rules are useless, nor that all scientific appraisals depend on metaphysics.

Thus far I have reported on those arguments of Feyerabend which had an *effect* on me. (He often tells us that arguments are to be evaluated in terms of their effectiveness, not their validity.) For completeness I should like now to refer briefly to a couple of points which Feyerabend thinks are of major importance but which left me unmoved or unconvinced.

First, the problem of the relationship between rules and practice. As I understand it, Feyerabend's position here is analogous to his account of theory and observation. Observation (*cf.* practice) is not just *laden* with theories (*cf.* rules); it is in part constituted by them. Conversely, theories (rules) cannot be understood in the abstract. They only gain meaning in conjunction with experience. I failed to see why this circumstance, even if true, would invalidate classical methodologies.

Secondly, there is the problem of comparing traditions. (It was in this context that the remark on racism *vs.* humanitarianism noted above came in.) His main point here is that it is question-begging to evaluate a tradition using the norms of that tradition. Somehow we have to go outside and use some other framework or tradition. Hence, he says, there can be no *objective* comparisons of traditions—'. . . racism will appear vicious to a humanitarian while humanitarianism will appear vapid to a racist' (p. 9). The tradition which values science and rationality is simply one among many—it has no objective (*i.e.*, framework-independent) claims to superiority.

Feyerabend's formal point seems to me to be correct. One discusses an object language in a meta-language; one uses meta-values in order to appraise value systems. There is no undogmatic way to stop an infinite regress and no way to justify one's starting point. But we can always criticise these starting points and in practice this situation need not lead to a breakdown in communication. Holders of incommensurable scientific theories *do* sometimes find or create a framework in which they can talk. And in my experience fruitful dialogue with sexists, racists, homophobes, fundamentalists, sociobiologists and punk rock fans is sometimes possible.

Of course, it is logically possible that people in opposing traditions share no basic values. But given the shared problem of survival which we all face, I think a total incommensurability of values is unlikely on evolutionary grounds. Surely the best strategy is to assume there is *some* minimal uniformity in people's moral and cognitive systems. (Of course, unlike Feyerabend, I do not live in California.)

---

[1] N. Koertge [1978].

2 SCIENCE IN A FREE SOCIETY

In the second section of his book, Feyerabend extends his argument into the area of political philosophy. In a democracy, all traditions or forms of life must be permitted as long as they do not seriously infringe on other people. However, in modern Western society, one tradition, that of scientific rationalism, has gotten out of hand and now threatens democracy. Since scientific rationality is just one ideology among many, it should not be given pride of place in any of our institutions, especially the schools. Power should be taken away from the scientific expert and given back to the people (who are then free to consult the expert if they wish).

I find Feyerabend's critique too wide-sweeping to be of much use (*e.g.*, in which areas do scientists really exert much power?) and his specific proposals for counteracting scientism and the cult of the expert rather naive. I agree that the medical establishment exploits and mystifies lay people horribly. But is a proliferation of acupuncturists and faith healers going to solve the problem? I agree that in America the melting pot tradition and general WASP arrogance has needlessly destroyed much that was valuable in the culture of minority groups. But should minority students now be *encouraged* to study Hopi medicine or Black dialects?

Feyerabend rightly realises that educational institutions play an enormous role in transmitting culture and it is here that his attack is most vehement. 'If the taxpayers of California want their state universities to teach Voodoo . . . and rain dance ceremonies, then this is what the universities will have to teach' (p. 87). Feyerabend does not mention the fact that in America local school boards and state legislatures already have a good deal of control over curriculum content and often act in ways which I believe Feyerabend would join me in condemning. In Texas, for example, standard dictionaries are now banned in the public schools because they contain secondary meanings for words such as *brain* (used as a transitive verb) and *bed* (as in 'go to bed with') which are considered to be too violent or salacious for high school students. So much the worse for the Texans he might say—they will mature by making mistakes. Unfortunately these seemingly local decisions have direct effects elsewhere. The Texas textbook market is so large that it could easily lead to nationwide revisions in school dictionaries.

Perhaps even more important than the issue of *who* makes such decisions is the question of *how* the decision making is to be institutionalised. In arguing for lay direction of all sectors of society (including the funding of scientific research), Feyerabend points out that lay juries are routinely asked to deal with expert evidence. Although we have few empirical studies on how jury decisions are actually made, I think the jury model might be a good one to emulate. Notice, however, that a trial has several features which are not characteristic of most political debates (or philosophical ones, for that manner!) and which on the face of it seem quite unFeyerabendian: Both sides are given an equal opportunity to present their case. Witnesses are cross-examined and must answer straightforwardly. (No rhetoric or fogging is permitted.) There are strict rules concerning the admissibility of evidence (material which is likely to prejudice the jury is generally not allowed). Jurors are kept in seclusion—there is no opportunity for them to read propaganda or to be acted on by pressure groups. Although not perfect (*e.g.*, barristers *are* permitted to use persuasive arguments), the jury trial is intended to provide a forum in which *rational* argument can prevail. It is as about as far removed from the anarchistic verbal slug-fests which Feyerabend fondly recalls in his autobiography as anything I can think of. However, if something like the jury system is what Feyerabend has in mind when he calls for more democracy, then I am all for it.

NORETTA KOERTGE
History and Philosophy of Science
Indiana University

REFERENCES

KOERTGE, N. [1970]: 'Inter-theoretic Criticism and the Growth of Science', in R. C. Buck and R. S. Cohen (*eds.*): *Boston Studies in the Philosophy of Science*, **8**, Reidel. Pp. 160–73.

KOERTGE, N. [1978]: 'In Praise of Truth and Scientific Rationality: Comments on Laudan's *Progress and Its Problems*' in *PSA*, **2**, forthcoming from the Philosophy of Science Association.

GREGORY, FREDERICK [1977]: *Scientific Materialism in Nineteenth Century Germany*, Studies in the History of Modern Science, vol. 1. Dordrecht: D. Reidel. Hardback $28, Paperback $13.50. Pp. xxiii+279.

1 Philosophical thought in Germany in the nineteenth century oscillated between the poles of naturalism and anti-naturalism. At the beginning the anti-naturalists held sway. Their most influential spokesman was, no doubt, Hegel, but the influence of Kant, Fichte and Schelling was also important in that period. The anti-naturalists tended to think that it was the role of philosophy to provide foundations for knowledge, including science, and that such knowledge could only be achieved by deduction from self-evident, *a priori* truths. They also believed that ultimate reality was spiritual, rather than material.

When Hegel was unexpectedly struck down by the cholera in 1831 his friends and disciples hoped that his philosophical system would live on. But things were to come differently. The social and political climate had shifted in Germany; empirical science and modern technology were gaining ascendance; materialist and empiricist ideas were infiltrating Germany from France and England. As a result German thought in the second third of the century turned towards naturalism. The natural sciences succeeded in throwing off the yoke of philosophical speculation and the empirical and experimental method was firmly established as the proper method of scientific investigation. The physiologist

Johannes Müller and his school (to which H. v. Helmholtz belonged) contributed significantly to this development. Müller and his students also insisted that the study of human psychology needed to be established on physiological principles. In philosophy itself the attack on Hegelian idealism had begun with Feuerbach's *Essence of Christianity* of 1841. Feuerbach's thought, in turn, gave rise to two different forms of materialism, the dialectical materialism of Marx and Engels and the scientific materialism of Vogt, Büchner, Moleschott and Czolbe.

The effect of these developments on academic philosophy was curious. Many of the new naturalists were scientists rather than philosophers and considered their philosophical writings mere extrapolations from their scientific research (*e.g.* Helmholtz). Others, such as the dialectical and scientific materialists, were excluded from academic positions for political reasons. Academic philosophers meanwhile continued with their idealistic constructions or turned to historical scholarship. As a result general interest in academic philosophy underwent rapid decline. Where Hegel and Schelling had lectured to huge audiences and had influenced with their rhetoric not only the course of philosophy, but also that of politics, science, poetry and theology, the academic philosophers of the mid-century contributed nothing to the culture at large. Materialism and, more generally, naturalism became the dominant ideologies of the age.

But this dominance was not to last. In the final third of the century epistemological considerations about the status of the scientific theories on which the materialists had built their claims stimulated a renewed interest in Kantian thought. The failure of the materialists to come up with a convincing account of mental phenomena led to the rise of anti-reductionist philosophies of mind, to monist philosophies modelled on Spinoza's thought, and to a renewal of classical empiricism (or subjective idealism) with its claim to be able to construct the world out of ideas or sensations. In addition, new developments in mathematics, in particular the work of Cantor, Dedekind and Frege, provided the impetus for platonistic and rationalist considerations.

Nevertheless, the naturalism of the middle of the century left important traces in the subsequent period. The neo-Kantians, who constituted the largest of the new schools of philosophy, respected the autonomy of the natural sciences and unlike the earlier idealists did not claim to be able to reduce science to *a priori* principles. They were content with assuring an area of knowledge that could not be reduced to empirical science. Unlike the Hegelians and unlike the mid-century materialists they preached a methodological dualism. The new developments in mathematics and logic and the impossibility of the reduction of these sciences to psychology seemed to accord well with this position. Some of the other new movements in philosophy were even closer to mid-century naturalism. Haeckel's monism rejected the mechanistic account of the mind given by the scientific materialists, but re-affirmed their belief that philosophy was ready to merge into empirical science. The critical positivism of Avenarius and Mach also maintained the primacy of empirical knowledge. The effect of mid-century naturalism can finally also be seen in the writings of such individual thinkers as Nietzsche and Freud.

While the materialist phase in German thought was a passing one it exerted an influence on all the new movements in the subsequent period of German thought from which many of the contemporary ideas in philosophy are derived. It is this continuity of influence that gives the study of mid-nineteenth century materialism its significance.

2 Because of its current political implications dialectical materialism has continued to be the subject of historical and philosophical studies. The same is not true of the scientific materialism of Vogt, Büchner, Moleschott and Czolbe. Their form of materialism faded away quickly in the last quarter of the nineteenth century without any direct descendants. Consequently it has not received attention commensurate to its erstwhile influence.

Frederick Gregory has now set out to fill this gap. His book provides the first comprehensive account of the rise and fall of scientific materialism in nineteenth century Germany. The book, which is both thoroughly researched and readable, should be of interest not only to historians of philosophy but also to philosophers.

Gregory begins his account with a discussion of Feuerbach's critique of Hegelian thought. He rightly sees Feuerbach as the father of both dialectical and scientific materialism. In *The Essence of Christianity* Feuerbach had set out to criticize both established religion and speculative philosophy. Knowledge, he had claimed, cannot be acquired through *a priori* reasoning but only through sensory observations of concrete facts. Idealism has reversed the true relation of body and mind; the physical world is not a property of the mind, but the mind a property of the body. Man's God is man himself and not some transcendent being. Religion must become anthropological, its concern must be the welfare of human beings on this earth. The implications of Feuerbach's philosophy were thus anti-speculative, empiricist, materialist, anthropological and politically radical. Gregory argues that the scientific materialists accepted all these implications of Feuerbach's thought and made them the starting-points of their own reasoning.

If there is one thing in which the four protagonists of scientific materialism differed from Feuerbach it was in their educational background and the direction of their interest. Feuerbach had begun his reflections on the basis of a study of Christianity and Hegelian philosophy. In four individual chapters on the scientific materialists Gregory describes their training and development and shows how each one of them received his initial training in physiology and medicine. Their knowledge of philosophy and their interest in it was limited. Moleschott, who had some acquaintance with philosophy, believed that philosophy and science were about to merge into one. Others were convinced that empirical science was making philosophy altogether superfluous (p. 146). Of the leading figures of the movement it was only Heinrich Czolbe (characteristically, the socially and politically most conservative of the scientific materialists) who thought it necessary to work out the philosophical implications of scientific materialism.

In accordance with their anti-speculative attitude the scientific materialists saw themselves as practical men rather than as theoreticians. The study of Feuerbach and physiology had convinced them of the truth of materialism and they considered that truth to have important practical significance. Their first goal was therefore to spread the acceptance of materialism through popular writings, to combat ignorance and superstition, to defeat the forces of clericalism and reaction. Characteristic is Jacob Moleschott's *Die Lehre der Nahrungsmittel: Für das Volk* (translated into English as *The Chemistry of Food and Diet*) of 1850 in which the author set out the requirements of a balanced diet and gave advice on how to prepare and cook foods. Only through improved diet could the social and political condition of the disadvantaged be improved. "Without phosphorous, no thought", as Moleschott said in his book (p. 91). Or as Feuerbach put it in his highly laudatory review of the book: "*Der Mensch ist, was er isst.*"

With their practical interest in materialism the scientific materialists were similar to Marx and Engels. The two groups shared a hostility towards the established powers and, in particular, towards established religion. Gregory sums up his book with the assertion that 'the overwhelming trademark of the scientific materialist... is not their materialism, but their atheism, more properly their humanistic religion' (p. 213). The scientific materialists differed from Marx and Engels however in not accepting the Hegelian notion of dialectic. They could therefore not agree with the dialectical materialists that historical change could only come through dialectical, revolutionary steps. The scientific materialists tended to be politically evolutionary, rather than revolutionary. Their views were closely connected with the radical liberalism of the 1848 revolution. While they were somewhat sympathetic to socialist ideas, they opposed the use of force to gain political advantage. In the end Büchner went so far as to favour a modified form of capitalism (p. 207). Such reformism drew upon the scientific materialists the furor of Marx and his followers.

Philosophically the scientific materialists believed in the unity of science and philosophy. They were completely opposed to the methodological dualism of the neo-Kantians and they had no sympathy for the idealism and apriorism inherent in Kant's philosophy. Scientifically, they argued against vitalism in biology which made its first appearance in the middle of the century. While they welcomed Darwin's work, their political progressivism inclined them to Lamarckism. (Czolbe on the other hand went so far as to argue against any evolutionary theory on general materialist grounds.) As far as the body-mind problem was concerned the scientific materialists tended towards a straight-forward mechanistic reductionism.

3 If one is to fault Gregory's book for anything it is for its failure to give a philosophically coherent account of the shortcomings of scientific materialism. The reasons for the decline of scientific materialism in the last third of the nineteenth century thus remain obscure.

I think that there were three major reasons. First, the scientific materialists had a rather naive view of scientific theories such as the physiological theories on which they based their conclusions. Second, they were not really very successful in accounting for mental phenomena in physiological and material terms. Third, they paid insufficient attention to logic and mathematics and their attempts to explain logical and mathematical principles in psychological terms were not well thought out.

The scientific materialists believed that scientific theories such as theories concerning human metabolism or Mayer's principle of the conservation of energy could be used directly to refute metaphysical doctrines such as the belief in the existence of God or the assumption of mind-body dualism. It never occurred to them that the relation of science to metaphysics might be more complex. They thought of the two enterprises as straightforwardly competitive and, hence, as either in agreement with each other or contradicting one another.

We can see why this view is unsatisfactory by considering the mind–body problem. The problem arises because at the descriptive level we can distinguish between mental and material (physiological) phenomena. The materialist thesis holds that every mental phenomenon will eventually be accountable for in material terms, whereas the dualist will deny this. Let us consider the case most favourable to the materialist. We assume that for every descriptively mental phenomenon we possess a corresponding account in material terms. Even in this case dualism is not decisively refuted. For, as the current discussion of the identity theory shows, it is always open to the dualist to claim that the materialist account establishes merely a correlation between mental and material phenomena, but that the phenomena are still distinct. Let us next consider the case favourable to the dualist. We assume that, however much our knowledge increases, we cannot find a materialist account for every descriptively mental phenomenon. Even in this case materialism is not refuted. For it is open to the materialist to argue that it is only the limitations of human understanding which stop us from providing the appropriate material account. Dualism and materialism are, thus, not testable theories and they can therefore not be said to be either proven or disproven by testable scientific theories.

At least one of the scientific materialists seems to have understood this point. In his book *Neue Darstellung des Sensualismus* Heinrich Czolbe argues that the materialistic assumptions cannot be proved and that, like the assumption of dualism, they are conventional and that therefore both materialism and dualism are logically equally justified. One can therefore call the principle of sensualistic materialism a prejudice or a preconceived opinion. Czolbe writes:

> Without such a prejudice the formation of any view concerning the connection of phenomena is altogether impossible. When scientists think they can draw concepts, judgments, and inferences out of sensory observation without any preconceived opinion, their view is likely to rest on self-deception. (*Neue Darstellung*, p. 6)

What recommends materialism to Czolbe is that it can exclude nonsensory entities; but in order to show this, he believes, one must actually prove that mental phenomena can be described in material terms. The other scientific materialists never made such an attempt. They were satisfied with the view that empirical theories could be derived from observation and that they could be used to refute metaphysical assumptions such as that of mind–body dualism.

Like Feuerbach the scientific materialists considered themselves both realists and empiricists. They did not, at least initially, see the difficulty in maintaining both these positions. They considered their materialism justified on the basis of a restriction to what is given through the senses. As the philosophical debate proceeded it became clearer that there was indeed a tension between the realism and the empiricism. If one was to grant only what is given by the senses, should one then not deny the existence of anything but sensations? Czolbe, as always more perceptive than his fellow materialists, eventually saw the difficulty. Its effect was to help in the destruction of materialism and to promote the kind of empiricism advocated by Mill. Subjective idealism and empirio-criticism began to take the place of realistic materialism.

Whereas the scientific materialists were inclined to consider theories true or false pictures of the world, the critical philosophers of the late nineteenth century had a more elaborate view of the role of scientific theorising. They realised that the mathematical, geometrical and other theoretical concepts that are used in scientific theorising are not in any way derivable from brute observation. Such considerations inspired a return to the Kantian idea that there are *a priori* forms of understanding which we must employ in order to form any coherent view of the world. On the other hand Mach and others stressed the instrumental function of scientific theories. The naive philosophy of science presupposed by the scientific materialists thus gave way to considerations out of which the contemporary discussion in the philosophy of science has emerged.

As far as the materialist analysis of mental phenomena is concerned the scientific materialists completely underestimated the difficulty of the project. One must not forget that their writings preceded the rise of experimental psychology. Their view of the mind was largely restricted to a belief in the association of ideas. Their views on neurophysiology were, of necessity, equally restricted. Czolbe's attempt to give a physiological account of mental process goes in principle not much beyond the hydraulic picture that Descartes had of the functioning of the nervous system. Contemporary materialism relies on notions of recursion and feedback in its account of mental phenomena. It models mental process in digital computers. It stresses the role language plays in intellectual processes. It allows for pattern recognition by the human sensory system. All such considerations are completely outside the scope of scientific materialism.

If we have learnt anything from the debate over the body–mind problem in the last hundred years, it is that even today materialism is, at best, a program, rather than a worked-out theory. Moreover, once we abandon the metaphysical aspirations of the past, it is really of little interest whether scientific theories lend support to the thesis that body and mind are identical or whether they show merely that there are correlations between mental and physiological phenomena. The dispute may come down to the question whether we prefer to call the bottle half full or half empty. What matters is to find out to what extent we can succeed in determining correlations between descriptively mental and descriptively physiological phenomena.

Finally, there is the failure of scientific materialism with respect to logic and mathematics. The tendency of the materialists was to account for the principles of logic and mathematics in terms of psychology. Czolbe went so far as to characterise them as empirical laws of the functioning of the physiological mechanism of the nervous system. In other words, the scientific materialists were thorough-going psychologists. This psychologism fails because it does not recognise that human reasoning need not follow the valid laws of logic. Such laws are, at best, normative; they do not describe actual human reasoning. Moreover, both the materialists' simple-minded view of the functioning of the human mind and their rather naive view of logic and mathematics stopped them from advancing either plausible or interesting hypotheses on the matter. In the final quarter of the nineteenth century Cantor and Frege advanced mathematical and logical theories which both of them regarded as incompatible with naturalistic assumptions. In a similar spirit the neo-Kantians insisted on the *a priori*, normative character of logical and mathematical principles. The scientific materialists had no reply to such arguments. Today one might argue that only a formalist view of logic and mathematics is fully reconcilable with the kind of materialism envisaged by them. But the scientific materialists neither saw this point nor would they have been able to work out a satisfactorily formalist position. For it was precisely the new work in logic and mathematics that proved the indispensible stepping-stone towards such a position.

In retrospect it seems that not much can be learned from the specific views of the scientific materialists. That is no reason to consider them inferior minds. Unlike the German idealists before them and unlike much of the philosophical tradition after them they were convinced that, in an age of science, philosophy would have to proceed in collaboration with science and not as a purely abstract, speculative enterprise. We cannot blame them for the fact that science has made enormous progress since the middle of the nineteenth century and that, from hindsight, many of the philosophical doctrines of the scientific materialists strike us as naive and outdated.

HANS D. SLUGA
University of California, Berkeley

HOOKWAY, C. and PETTIT, P. (*eds.*) [1978]: *Action and Interpretation*. Studies in the Philosophy of the Social Sciences. Cambridge: Cambridge University Press 1978. Pp. xii+178. £5·95.

Before reviewing this book I had better précis the papers in their order of appearance. 'Introduction', by the editors. The problems of the philosophy of the social sciences cluster round the explanation of social phenomena, choice of theories of social phenomena, and reduction of social phenomena to, *e.g.* physical or psychological phenomena. Only the first two topics are discussed in this volume. Two themes run through the papers: that intentions, beliefs and desires, being invisible, cannot be known about or attributed to with certainty; and, experience underdetermines theories ('attending to experience will not reduce the number of explanatory theories to one', p. x), so, principles of theory-choice have to be employed.

Mary Hesse, 'Theory and Value in the Social Sciences'. '... the backwardness, theoretical triviality, and empirical rule-of-thumb character of most social science' (p. 7) has to do with there being no reasonable expectation of social science ever being able to predict, *perhaps* because the social environment is such that no computer of limited capacity can learn to predict it. But the decision to assess theories by the pragmatic criterion of predictive success involves a value judgment and we are free to choose other criteria by which to assess the social sciences.

Christopher Hookway, 'Indeterminacy and Interpretation'. 'Quine's thesis of the indeterminacy of translation attempts to justify a non-realist attitude to psychological states by arguing from the problems of the translator ... [I]t imposes constraints upon . . . those attempting to understand other cultures. It provides . . . arguments that urge a radical discontinuity between the social and physical sciences' (p. 18). In summary, that argument goes:

> manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another.

(p. 21, quoting W. V. O. Quine, *Word and Object*, Cambridge (Mass.) 1960, p. 27). Hookway thinks this argument does not in the end vitiate anthropological practice but perhaps alerts us to the problems and options. An interpretation of other people's utterances that treats them as persons (and that move involves attributing a great deal) and hence able to converse, engage in and learn from cooperative inquiry, is to be preferred.

Philip Pettit, 'Rational Man Theory'. The presumption of rationality in others enables us to communicate and interact, *i.e.* partly to control others' expectations of us—however 'soft-edged' such a presumption makes the social sciences. Rational man theory allows us to see how to interact and communicate with people like delinquents and Milgram's sadists. In application, rational man theory is an art not a science (p. 60).

Alan Ryan, 'Maximising, Moralising and Dramatising'. A ruminative discussion of Erving Goffman, that does no more than ask if his type of dramaturgical explanation is *sui generis*, and whether it is possible to distinguish the medium from the message in the passing of information through symbols.

John Skorupski, 'The Meaning of Another Culture's Beliefs', distinguishes literalism (people mean just what they say) from intellectualism (everyday remarks and actions presuppose cosmologies), and contrasts both with symbolism, Winch-ism and relativism. He questions whether the latter 'can provide sufficient controls for assessing the correctness of a given account of the meaning of ritual discourse' (p. 105). And goes on:

> can what people in a given culture mean, on the given theory of meaning/principles of interpretation, by what they say, be linked with a plausible psycho-sociological theory which tells us why they should be given to saying what—so understood—they do say? When this question is properly posed, then the candidacy of that account of primitive cosmology which we have been considering—classical and realist in its semantic principles of interpretation, intellectualist in its sociology of thought—looks much stronger than some of the orthodoxies of modern social science and philosophy have allowed (p. 106).

Nick Jardine, '"Realistic" Realism and the Progress of Science'. History of science as a case study to defeat relativism. Jardine labours to defeat Putnam's principle of the benefit of the doubt which says that the properties and entities of past science can only be equated with the properties and entities of present science if the descriptions which the protagonists of past science used would when reasonably reformulated characterise present science. Finding Putnam's relativism unacceptable, Jardine offers in its stead principles to save observations and causal attributions as they are to be found in past science and which yet assimilate the relativist insight about the diversity of culture's ideas about nature.

The vision of science he offers is of an enterprise endlessly tenative and open-ended.

John McDowell, 'On "The Reality of the Past" '. 'The topic of this paper is not the reality of the past, but rather realism and anti-realism as discussed by Michael Dummett . . .' (p. 127 *n*), who considers how to get round the argument that 'linguistic competance cannot involve a conception of sorts of circumstance other than those which a language-learner had available to his consciousness in learning the language' (p. 129) and that apparently cannot include sentences about the past. To try to link present and past by declaring true such past-tensed sentences as correspond to suitable present-tense sentences that could have been truly uttered in the past does not escape the problem. Dummett thinks realism may have to be given up. McDowell thinks not. Realism can be formulated so as to distinguish how we learn to use language from the truth-conditions of that use.

John Dunn, 'Practising History and Social Science on "Realist" Assumptions'. If there are no methodological principles that can guarantee that we will come to know some of the truth about history and the social sciences is not their claim to have a subject-matter or be a field a 'bare-faced fraud' (p. 146)? Quine's indeterminacy thesis is a problem. 'We are all of us unsurprised in our own social life for a great deal of the time' (p. 155). Our duty is to do social science so as to render each man his due. This seems to require realism.

Every reader brings his own interpretation to a text, attempts to impose it, and attends closely to the passages that resist his interpretation. These points of resistance resemble refutations in science to the extent that where the text resists an interpretation it offers the reader the opportunity to learn from the experience of reading by modifying his interpretation. In the case of volumes by several hands this process is usually too complicated for the reader to report on it in detail, since he may bring a different interpretation to each paper. In the case of the present volume, however, there are similarities of reference, texture and style that would seem to simplify matters. Yet they have not. To my surprise, no interpretation I tried on the book fit nearly well enough to bother with points of resistance. This was chastening. I had thought myself fairly well-positioned to know what is going on in the philosophy of the social sciences, and hence how a volume like this was to be interpreted. Clearly the boundaries of the field itself ramify in ways of which I was shockingly unaware. In particular, philosophy of the social sciences now appears in danger of being assimilated to that branch of latter-day analytic language philosophy misleadingly called 'philosophy of mind'. Exposure to this turn of events has not enamoured me of them, but does alert me to the possibility I may have indulged in wilful avoidance. It is time to face things.

Appearances first. This is a Cambridge volume. Two of the symposiasts are at Oxford, though their roots may be in Cambridge. What does Cambridge signify on the map of philosophy nowadays? The university of Russell, Ramsey and Broad was also, one reminds oneself with a shudder, the university of Moore and Wittgenstein. So is it any wonder that the style of argumentation in this volume could be described as latter-day analytic philosophy with a very strong linguistic turn and with a technical-sounding ring to it? Less easy to understand is its relentless professionalism and dryness, besetting sins of so much current philosophical writing. ('Attending to experience will not reduce the number of explanatory theories to one', and stand up the boy who said it did!) There are two contributors who earn their living teaching politics but really you would hardly notice it from the way they write. The only social scientist given more than passing mention is Erving Goffman, an intriguing enough figure (see *Philosophy of the Social Sciences*, **8**, 1978, 67–87), but singling him out is rather like philosophers of the natural sciences discussing only the philosophy of Werner Heisenberg. Not only are actual social science examples eschewed, but so, I am afraid, are classical philosophers of the social sciences. Instead one notes that the most frequent names cited are Wittgenstein, Davidson, Quine, Dummett, Putnam and Lewis. The blurb informs us that the philosophy of the social sciences is a 'fashionable but rather diffuse area of interest', and important implications for it are to be found in 'recent developments in the philosophy of language'—developments associated, apparently, with these names. The list of names is anything but diffuse in interests and is certainly very fashionable. However, they are almost the last philosophers I should associate with the serious study of philosophical problems of the social sciences, as I shall explain.

How have the rather esoteric concerns of these fashionable philosophers of language come to appropriate space on library shelves devoted to philosophy of the social sciences? To begin with, there is fashion itself. Bright young men want to discuss the views of those they admire and, as Anthony Quinton informs us in an excellent recent sketch-map of philosophy as the official establishment currently sees it, Donald Davidson is 'the philosopher most admired by the youngest and liveliest generation' (*Times Higher Education Supplement*, no. 371, 30 November 1979, p. 9). But when those they admire produce perverse arguments their good sense makes them resist. They then mistakenly conclude that confounding these arguments is a worthwhile enterprise that derives its importance from the admired authors. They overlook the slide from admiration and fashion to importance, and neglect the possibility that such authors might have produced arguments that are not worth bothering about. Hence these young men are a bit like the original victims of language philosophy, which used regularly to expose students to old-fashioned philosophy in order to guarantee employment to those professionally engaged in dispelling conceptual confusion such as was to be found in old-fashioned philosophy . . .

But the sociology of thought needs to be supplemented with connections of ideas; what is the connection between language philosophy and the philosophy of the social sciences? Half the answer is that philosophical problems have been claimed to have their origin in language, and Wittgenstein and his followers somehow stumbled across the idea that language was a social institution and that the manner it performed could not be separated from that social context. This problem is seen in its most dramatic form when the social scientist is simultaneously trying to explain some social phenomenon and trying to translate it from an alien language into his own, *i.e.* anthropology. The other half of the answer is that social scientists find language a crucial and baffling social institution. Seemingly unique to man, not designed, highly functional, deeply related to the very possibility of all thought, including sociological thought about, among other things, language itself, endlessly discussed by those ambivalently admired/sneered-at purists philosophers, it demands attention. Yet the language philosophers cited in this book have scarcely discussed language as a social institution at all. What they have done, to judge by this book, is to manufacture some arguments concerning the very possibility of doing social science. From considerations to do with translation, manners of speaking and tensed utterances they move to difficulties in talking about the past (which almost all social science necessarily does), in treating the social world as really there, in studying other people and other cultures. Quine especially is taken, as Hookway makes clear, to have challenged the possibility of properly translating from another language and hence of coming by that means to understand another culture. Davidson has argued for a physicalist monism, and both Dummett and Putnam have in different ways raised doubts about whether realism can be consistently maintained. They form a cosy circle because to a considerable extent they debate mainly with each other. But why on earth should social scientists (as opposed to impressionable young philosophical professionals) take seriously newish sceptical arguments concerning the very possibility of what they are doing? The usual argument is that philosopher's questions are preliminary or foundational. In the heyday of both logical positivism and language philosophy the suggestion was that first-order scientific work would only be put on a sound footing when the foundational problems were solved. But scientific work could not wait on the completion of the International Encyclopaedia of Unified Science.

The kind of philosophising social scientists do take seriously emerges from difficulties in first-order work—think of Durkheim, Weber, Pareto, Robbins, Hayek, Popper, Hutchison. No doubt it is possible for work of interest to emerge from the concerns of pure philosophy but not, I think, when what is offered is sceptical, anti-realist arguments. The logic of the situation is as follows. A robust science has its problems, theories and debate. An argument, however ingenious, casting doubt on the whole enterprise, and substituting, let us say, a non-realistic account of other persons, other cultures or the past is unlikely to shake such a science even if no flaw can be detected in it. A simple common-sense realism seems to be a presupposition of serious science, and a moral condition for dealing with fellow human beings; moreover, the alternative is highly implausible. Realism is a conjecture that is held critically. This does not mean that when it is in difficulty we immediately give it up. Russell said of idealism that it was impossible to refute by argument, but then again there wasn't the slightest reason to believe that it was true. Much the same applies to Quine's indeterminacy thesis and the anti-realism that we hear about in this volume. To say of translation that it cannot be perfect is not to say that it cannot be undertaken with more or less success; to say that statements about the past or about the real world raise philosophical puzzles is not to say that should not be made, still less is it to say that we should all switch over to some non-realistic philosophy.

So what we have in this volume is some clever young philosophers who seem not to have heard that the idea that the social sciences must offer some form of defence of themselves in the face of sceptical objections is yet another version of outdated justificationist philosophy. We do not demand that physical science defend itself against arguments by sceptics, idealists, phenomenologists and language philosophers to the effect that it has insoluble foundational problems.

This may be because we all have some grasp of and respect for the problem-situation in the physical sciences. Why then do we not respect the problem-situation in the social sciences? Vigorous debate is in progress about theoretical and practical problems concerning the economy, social pathology, the workings of other cultures, what actually happened, and so on. Arguments suggesting their unreality, the problematic character of all talk about them, the uncertainty of meaning what you say are rightly going to be given short shrift.

I. C. JARVIE
York University, Toronto

SALMON, WESLEY C. (*ed.*) [1979]: *Hans Reichenbach: Logical Empiricist.* Dordrecht, Holland; Boston, U.S.A.; and London, England: D. Reidel Publishing Co. Pp. 782. $39.50.

There is no volume in *The Library of Living Philosophers* devoted to Reichenbach. A volume was planned on *Logical Empiricism* to be devoted jointly to Carnap and Reichenbach; but Reichenbach died in 1953 while the volume was in the early states of its preparation, and the volume in consequence was devoted entirely to Carnap. The present volume, a large collection of essays on all the various aspects of Reichenbach's work, is an attempt to make good this deficiency. Most of the articles in the volume were originally published in *Synthese* in 1977. The volume comes at a time when a nine-volume German edition of Reichenbach's *Collected Works* is being published; and a considerable two-volume English edition of Reichenbach's less well known papers *Hans Reichenbach: Selected Writings 1909–1953* has just been published. Reichenbach has indeed been very fortunate in the dedication of his disciple Wesley Salmon and of his second wife Maria Reichenbach. Over the past twenty seven years together they have secured the publication of his unpublished papers and books, the reprinting or publication in English of books which have gone out of print or were published only in German, and the widespread dissemination and discussion of his ideas. The philosophical community has also very good reason to be grateful for all this effort; for Reichenbach, one of the great empiricists who made the philosophy of science what it is today, had much of great value to say on all aspects of the philosophy of science.

Reichenbach sought to represent science as a rational activity concerned to predict new observations on the basis of old observations. It was concerned basically only with what was in principle observable. Although science might include statements about the unobservable—about the geometry of space, or subatomic waves and particles—such statements only had a normal truth-value when combined with other statements which carried consequences about possible observations. Science was a rational activity because the method of extrapolating from past observations by the "straight rule" could be shown to be at least as likely to yield success as any other method. The universal statements obtained by this method (complicated by the application of Bayes' Theorem in a state of "advanced knowledge") were nomological statements. Nomological statements were to be distinguished from other universal statements basically by the method of their justification (*i.e.* ultimately in terms of the kind of observations which were evidence for them); they did not report any unanalysable "physical necessity". Salmon—surely correctly—sees Reichenbach's work as a reaction against the Kantian tradition, and its dominating theme as the reconstruction of (p. 2) 'knowledge in general—and scientific knowledge in particular—without recourse to the synthetic *a priori* in any form'.

The many contributors to this volume have varying degrees of sympathy for Reichenbach's programme. Wesley Salmon remains, as one would expect, basically highly sympathetic to it, concerned only to refine the details of Reichenbach's analyses. The volume begins with a helpful eighty-page introduction by Salmon, outlining Reichenbach's main contributions to philosophy in a clear and elementary way. There follows next a paper by F. John Clendinnen on Reichenbach's "pragmatic" justification of induction. He makes a number of criticisms, some of them common ones. Thus Reichenbach claims that if any method of prediction would have been successful in some field, so would normal scientific induction. His point is that if there is another method which is uniformly successful, the scientist could have noted the correlation between the predictions of that method and what actually happened, and he could then have generalised this to give a generalisation (*e.g.* "all the predictions of clairvoyant C are true") which would yield successful predictions. Clendinnen imagines a world where one clairvoyant C is successful up to t_2 but not thereafter, whereas clairvoyant D is successful after t_2. There is a method of prediction which would have been uniformly successful—*viz.* "follow C up to t_2 and thereafter follow D". Yet scientific induction would never have been able to obtain a generalisation which would give successful predictions until well after t_2. More generally over any finite time there may be a method of prediction which is more successful than scientific induction. Clendinnen argues, to my mind correctly, that the general principle of induction is (p. 114) "to seek the simplest overall systematisation of knowledge". Prior probabilities must be assigned to hypotheses on the basis of their simplicity and not solely, as Reichenbach and Salmon hold, on the basis of frequency considerations. Clendinnen goes on to claim that 'it is only by adopting the simplest generalisation compatible with the known facts that we are guided by the facts themselves'. This is a valuable point, but Clendinnen seems to regard it as the basis for a vindication of induction. It seems to me however still to leave wide open the question of what reason we have for supposing that our predictions will be successful if we are "guided by the facts".

Next, there are three papers on Reichenbach's frequency theory of probability. Von Fraassen is concerned with reforming Reichenbach's mathematical apparatus; Ben Rogers with the problem of construing the probabilities of theories in terms of frequencies; and James H. Fetzer with the problem of assigning probabilities to the occurrence of particular events. The latter problem is of course that particular events can be classified as belonging to many different reference classes; yet, according to how they are classified, so on a frequency theory the probability of their occurrence will differ. Fetzer brings out in a very competent way the difficulties in attempts by frequency theorists to specify the class to which we should refer particular events, *e.g.* Salmon's solution that we should refer them to the broadest epistemically homogeneous reference class to which they belong.

Then there are a number of papers on Reichenbach's philosophy of relativity and geometry. These include a technical piece of physics (Adolf Grünbaum and Allen I. Janis on 'The Geometry of the Rotating Disk in the Special Theory of Relativity') and two historical papers, one by Andreas Kamlah which urges us to understand Reichenbach's "conventionalism" against the background of Kant's thought-orientated semantics, and one by J. Alberto Coffa which compares Reichenbach's theory of geometry with that of Weyl. Keith Lehrer, Clark Glymour and Laurent A. Beauregard investigate the kind of conventionalism about geometry and simultaneity which Reichenbach was advocating, and also to some extent whether his fact/convention distinction is a satisfactory one (though Lehrer seems to take for granted that basically the distinction is satisfactory).

Generally I found the papers on Reichenbach's work on Quantum Theory more illuminating than those on Relativity Theory. Donald Richard Nilson has a long, lucid and illuminating commentary on Reichenbach's Quantum Logic and the criticisms which have been made of it. Gary Hardegree and Roger Jones have more technical papers. Neal Grossman makes a powerful plea to philosophers to take seriously the metaphysical implications of Quantum Theory, as brought out by Reichenbach's Principle of Anomaly. This principle claims that any exhaustive interpretation of the world of Quantum Mechanics (*viz.* one which fills in the gaps between the 'phenomena' of particle-interactions) involves causal anomalies. For this reason Reichenbach preferred a restrictive interpretation, that is one which gave the truth-values 'true' or 'false' only to statements about phenomena. But Grossman argues that such an interpretation fails to explain the phenomena; to provide explanation we need statements about what Reichenbach called 'interphenomena'. But we cannot regard the interphenomena either as states of waves or as states of particles. So we must abandon atomistic explanation altogether, and postulate that the basic unobservable entities whose behaviour causes the phenomena are of a more mysterious kind and operate holistically.

There are papers by O. Costa de Beauregard, Laurence Sklar and Philip von Bretzel on the area of "the direction of time". Costa de Beauregard hardly mentions Reichenbach and is very speculative. Sklar by contrast is very hesitant and careful. His discussion of the difficulties of deriving the topology of space-time from relations of causal connectability is perhaps the most philosophically sensitive piece in the volume. Von Bretzel develops Reichenbach's idea that, as coincidences have a common cause but not a common effect, we can distinguish causes from effects independently of temporal considerations; and thus recognise temporal order by causal order. He wisely construes the claim (p. 385) 'that the temporal order of the events of the universe is given by the causal order' as 'not a claim about the meanings of terms but a theory about the nature of time.' Salmon's paper 'Why ask "Why?"' also really falls within this group. He asks why we seek for scientific explanations, and he answers that they reveal the mechanisms of the bringing about of events, and he too claims here that causes are to be distinguished from effects by the 'coincidence' test mentioned above.

Reichenbach's account of nomological statements is little known. His monograph on the subject, originally published posthumously in 1954 and reissued in 1976 is a very complicated piece of analysis. Salmon's introduction to the

1976 edition is republished in this volume. It is a helpful guide to that work and provides justification for Salmon's claim that Reichenbach's treatment of the topic has many valuable insights for philosophers. Evan Jobe also finds Reichenbach's theory of nomological statements of great value. He concludes (p. 719) that, 'it does seem to be successful, when appropriately revised, in demarcating the class of laws that are highly confirmed at some time or other', but not in demarcating the class of laws of nature.

No aspect of Reichenbach's work is neglected. There is a nice essay by Merilee H. Salmon on Reichenbach's views on the need for and justification of consistency proofs for applied mathematics. There is a paper by William E. McMahon on Reichenbach's project for creating a symbolic model for ordinary language, and a paper by Cynthia Schuster on Reichenbach's meta-ethics.

This is a rich volume full of valuable detailed exposition, criticism and development of Reichenbach's views. It brings vividly before us the range of his interests. Reichenbach worked out in careful detail his empiricist view of the world. He made it a plausible and attractive view. For myself, I think that this overall view was largely mistaken. I think that science is concerned to describe the unobservable reality which explains what we observe; that there is a physical necessity in nature; and that we approach nature with certain *a priori* principles of probability for which we can provide no further justification beyond the fact that they seem intuitively obvious. But this difference of overall view does not lead me to find the details of Reichenbach's work lacking in value. For it is easy to translate a lot of Reichenbach's detailed work into a different idiom—*e.g.* to regard his work on physical necessity as showing not the nature of physical necessity but the *a priori* criteria by which we judge that something is physically necessary; his work on simultaneity not as showing that some distant events are not "absolutely" "simultaneous" or "non-simultaneous" with each other, but as showing only the impossibility of determining whether or not some distant events are simultaneous. I think that Reichenbach's concentration on the details of modern physics distinguishes him from his other great contemporaries, Popper, Carnap, and Hempel; and this will allow much of his work to survive, even if his overall philosophical view becomes no longer a fashionable one.

RICHARD SWINBURNE
University of Keele

HUFF, D. and PREWETT, O. (*eds.*) [1979]: *The Nature of the Physical Universe: 1976 Nobel Conference*. New York: John Wiley and Sons. Pp. xvii+140. £12.75.

The collection of papers published in this volume originated as lectures delivered at the 1976 Nobel Conference organised by Gustavus Adolphus College in Minnesota in association with the Nobel Foundation in Sweden. As explained in the editorial preface the object of these conferences is to enable scientists to meet and discuss leading topics in science for the benefit of an audience of intelligent laymen. In the present instance there are more specifically philosophical contributions from Stanley Jaki and Hilary Putnam.

The first three papers give an elementary and essentially 'popular' account of particle physics. Victor Weisskopf in his paper entitled 'What is an Elementary Particle?' gives a brief but lucid history of how the notion of elementary particle has been gradually refined as we proceed along the sequence from molecules to atoms to atomic nuclei to nucleons to quarks. As we go to higher and higher energies of possible excitation successive layers of structure are revealed and Weisskopf concludes that we should really embrace a notion of *conditional elementarity* in which the objects we refer to as elementary depend on the energies involved in probing their structure. Weisskopf appears to subscribe to an infinite regress, but in his contribution Murray Gell-Mann asks the question 'What are the Building Blocks of Matter?' and concludes that the picture that has emerged during the last few years of two basic families of fundamental particle, the leptons on the one hand (electrons, muons and neutrinos) and quarks on the other (which come in a variety of 'flavours' and 'colours') together with the particles which mediate their interactions (such as photons, W-particles and gluons) may represent a truly fundamental level. Gell-Mann believes that Nature may have chosen 'some very elegant mathematical scheme' that requires just this particular number of particles.

One of the difficulties of writing about this sort of physics and relying on stop-press information is that everything is changing so quickly. Already the account given by Gell-Mann is essentially out-of-date in its details with the discovery of the *tau* lepton and with two new varieties of quark ('bottom' and 'top') inviting the attention of the theorists.

In the third paper Steven Weinberg asks 'Is Nature Simple?' and gives a delightful introduction to the notion of broken symmetry in particle physics. This is the idea that the basic laws of physics are very simple in the sense of exhibiting a high degree of symmetry, but the particular manifestations of these laws which we interpret as elementary particle systems may lack the symmetry of the underlying laws, thus exhibiting the phenomenon of spontaneous symmetry breaking, already familiar in other branches of physics such as solid-state physics. Weinberg explains in particular the essential rôle of gauge symmetries in modern theories of particle interactions. Here the possibility of performing different symmetry operations at different space-time points is exploited to constrain possible types of interaction. Expressed mathematically the gauge fields arise as infinitesimal connections on a fibre bundle whose base space is space-time and whose structural group is the dynamical symmetry in question. Weinberg confines himself to a purely non-mathematical exposition, which is nevertheless very well done.

Of course one could not ask for more prestigious guides to modern particle physics than Weisskopf, Gell-Mann and Weinberg (the last two are both Nobel Laureates). But when it comes to any sort of philosophical discussion everything is very superficial. The whole subject of modern particle physics cries out for detailed philosophical analysis. But the philosophers who could do the job do not know the physics while the physicists remain incredibly naive philosophically speaking. To take a couple of examples at random Gell-Mann refers to causality as 'a very simple principle . . . which states that a cause precedes its effect'. One is left completely in the dark as to whether he subscribes to a causal theory of time or a temporal theory of cause! Again Weinberg equates simplicity with symmetry, but every equation expressing a law of physics possesses as many symmetries as there are automorphisms inducible among its solutions. To get a handle on simplicity we must already understand which automorphisms are to be counted as simple. So the argument really goes in a circle.

Sir Fred Hoyle's paper entitled 'An Astronomer's View of the Evolution of Man' is largely concerned with the origin and limits of the Universe's material resources. Hoyle explains the modern views on how the various chemical elements are produced at different stages of the evolution of stars and how these elements are spread around in interstellar gas clouds by supernova explosions. Hoyle goes on to argue the plausibility of an interstellar origin for life which he considers may be very abundant on the cosmic scale. He stresses that our earthly chemical environment is cosmically determined in respect of the relative abundancies of various materials and then proceeds to draw the conclusion that only cosmically abundant elements can be useful to man in the long run except for these employed in high technology where only small quantities are required or if they are the source of nuclear power where low abundance may be compensated by very high yield in energy terms. In their introduction the editors claim that Hoyle argues that 'nuclear energy, and not solar energy, is the most preferable alternative to resolving the current energy crisis', a point which is reiterated, presumably on account of its topical relevance, in the publisher's blurb. However Hoyle does not in fact refer directly to the utilisation of solar energy. All he says is that nuclear energy is 'the only adequate replacement (for the hydrocarbons) presently known to be possible'. The editors quite wrongly impute to Hoyle the view that nuclear power is not just possible but also preferable as the source of our energy requirements. Hoyle concludes by arguing the vital importance of limiting the world's population. If birth and death rates were equalised 'nothing concerning the future of our species would then look unfavourable'. In an age of nuclear confrontation I wish I could share Hoyle's optimism.

I turn now to a rather curious contribution from Stanley Jaki entitled 'The Chaos of Scientific Cosmology'. Jaki here intends a play on the word 'chaos' taken either as referring to the primeval state of the universe or to what he regards as the chaotic state of the past and indeed present understanding of cosmology. Distinguished as theologian, philosopher, physicist and historian Jaki's piece is full of obscure erudition, but there are a number of detailed claims I find puzzling. For example Jaki argues that geometries of non-zero curvature tell strongly against an *a priori* approach to cosmology on the grounds that the particular value of the curvature must be an empirical matter. But it seems to me one could just as well have an *a priori* justification for a particular relationship between the curvature of the universe and other constants of nature as for justifying the Euclidean value of zero on *a priori* grounds. Indeed this was the point of view of Eddington. Then Jaki talks of the mischievous notion of infinity in physics, but the example he chooses is an unfortunate one. He refers to the 'infinity catastrophe' in the classical theory of black-body radiation as being corrected by Planck's notion of the discrete quantum of action, but while in one sense Planck abolished the infinitely small in this context, in another sense infinity is still retained with an infinite integral replaced by an infinite sum for the mean energy associated with vibrations of a particular frequency. A little later in the paper Jaki makes a plea that cosmologists should take serious note of Gödel's incompleteness theorems. Jaki is here calling in question the consistency of what he refers to as the 'complex . . . mathematical underpinnings of cosmological theory'. But I doubt if many mathematicians seriously believe that first-order Peano arithmetic or Zermelo–Fraenkel set theory may be inconsistent on the basis of Gödel's second incompleteness theorem. Though the consistency cannot be proved 'internally' there is 'external' evidence, not least the failure of mathematicians over a lengthy period actually to derive a contradiction. Perhaps Jaki thinks that some deep theorem in mathematical cosmology is a Gödelian undecidable sentence, but in a sense (admittedly not countenanced by the strict constructivist) Gödelian sentences are known to be true, the argument establishing this being conducted in the semantical rather than the syntactical metalanguage. I do not think that our ability to understand the cosmos is seriously threatened by Gödel's limitative results, their importance is rather to undermine a purely formalist approach to the philosophy of mathematics.

The final paper in the volume is by Hilary Putnam with the title 'The Place of Facts in a World of Values'. In this essay Putnam contends that there are 'facts' which lie quite outside the scope of empirical science. In particular Putnam asserts, but does not really argue for, the objective status of moral values. He also believes the freedom of the human will to be a factual truth. Putnam sees the indeterminacy of quantum mechanics as relevant here, not in the sense that human actions are caused by chance, far from it, but apparently because a causal hiatus in physical laws leaves room for the intervention of free-will. Of course one would be quite wrong to argue that quantum mechanics actually licences such intervention although there has been a regrettable tendency for some physicists, such as Wigner, to see the interaction of human consciousness and physical systems as an essential ingredient in understanding some of the paradoxical aspects of the theory of measurement in quantum mechanics. Finally Putnam turns to reason and rationality. What is reasonable, what is plausible, are matters of (non-scientific) fact. How do we account for the success of science? For Putnam any trial and error approach to the development of scientific theories is quite unsatisfactory in answering this question. Putnam himself inclines to a Kantian type of solution. The success of science is a puzzle 'as long as we view concepts and objects as radically independent; that is, as long as we think of "the world" as an entity that has a fixed nature, determined once and for all, independently of our framework of concepts'. Putnam concludes by affirming that science is arrived at by plausible reasoning which 'comes up with truths and approximate truths far more often than any trial and error procedure could be expected to do . . . common sense and critical intelligence . . . have to be brought *to* scientific . . . ideas; they are not a commodity to be purchased *from* "science".' Putnam admits that the main conclusion of his paper is not defended or even adequately expounded on this occasion and leaves the reader rather tantalizingly with reference to more detailed discussion elsewhere.

So on balance this book is not to be recommended for serious study of the scientific or philosophical issues expounded. Most of the essays have a pretty informal, after-dinner speechifying ring about them. The contributions from Weisskopf and Gell-Mann tend to patronise the reader, only Weinberg really succeeds in communicating and sharing his own excitement and enthusiasm at what he is talking about. My most obvious disappointment is the lack of any real meeting ground between the physicists and the philosophers. An opportunity for a really fruitful collaboration has been allowed to slip away and degenerate into a series of lightweight disjoint contributions that leaves the reader strangely unenlightened as to the essential entanglement of scientific theorising and philosophical analysis that must characterise any deep understanding of the nature of the physical universe.

MICHAEL REDHEAD
Chelsea College, University of London

Craig, W. L. [1979]: *The Kalām Cosmological Argument.* London: Library of Philosophy and Religion, Macmillan. Pp. 216. £12.00.

Dr Craig's object in this book is to defend one of the oldest and most controversial proofs of the existence of a Creator of the universe. It stems from the sixth-century Alexandrian philosophical commentator and Christian theologian Joannes Philoponos, who argued against Aristotle's proofs for the eternity of the universe, and was more fully developed by medieval Islamic theologians of the school to which the Arabic term *kalām*, signifying natural theology or philosophical deism, became attached. The argument was transmitted to the Latin West in the thirteenth century and became a source of controversy, being supported by Bonaventure and opposed by Aquinas.

Because modern western writers have usually been unaware that the argument was inherited from Muslim theologians and philosophers, Craig devotes the first part of his book to a scholarly historical account of the discussions by three of its most important Muslim defenders, al-Kindi in the ninth century, Saadia in the tenth century, and al-Ghāzāli (al-Tūsi) in the twelfth century. The second, and longer, part of the book is devoted to the author's own defence of the argument from a modern point of view. The book concludes with two appendices relating the argument to Zeno's paradoxes and the thesis of Kant's first antinomy, respectively.

There are two stages in the author's argument: in one the object is to show that the universe must have begun to exist at some definite time in the past, and in the other that this beginning must have been caused. Since much of the book is devoted to the former, its publication will be of particular interest to those who have followed the debate on the concept of an infinite past that has recently been conducted in the pages of this *Journal.*

Dr Craig, who is at present engaged in research at the University of Munich, obtained a Ph.D. in philosophy at the University of Birmingham, where his supervisor was the distinguished theologian John Hick who edits the series in which this book appears. In the course of preparing it he has made a careful study not only of the history of the *kalām* cosmological argument but also of the theory of infinite sets and of modern cosmology.

Craig develops two philosophical arguments in support of the view that the universe must necessarily have a finite past. The first is the general one that there cannot be ***any*** actual infinites in the physical world, which would consequently rule out the possibility of an infinite temporal regress of events. Craig's belief that the concept of the actual infinite of modern set theory is restricted to the abstract world of pure mathematics has the support of some of the leading experts. He points out that, according to Abraham Robinson, 'Cantor's infinites are abstract and divorced from the physical world', and David Hilbert, who so enthusiastically defended Cantor's theory against the objections of Brouwer, maintained that 'the infinite is nowhere to be found in reality'. Craig describes with considerable wit some of the apparent absurdities which the real existence of the actual infinite would entail. Moreover, although he points out that his case against the physical existence of the actual infinite, 'says nothing about the use of the idea of the infinite in conceptual mathematical systems', he reminds the reader that Cantor's theory has not been accepted by all mathematicians and has been rejected by the intuitionists.

A more direct, and therefore stronger, argument for the necessarily finite nature of the past is provided by Craig's second argument, which is specifically concerned with infinite temporal sequences. For, even if his first argument is rejected (or regarded as plausible rather than compelling) and the possibility of the universe being infinitely extended in space and material content is not definitely ruled out, it could still be argued that the past must necessarily be finite. I believe that this conclusion, which is strongly supported by Craig's second argument, is valid despite the recent criticism to which it has been subjected in this *Journal* by John Bell in an article that appeared too late for Craig to take into account (*B.J.P.S.*, 1979, xxx, 161–5).

Bell, like Russell and others, bases his case for the possibility of an infinite past on Cantor's theory of infinite sets. His article is a good example of the confusion that can result from assuming that this abstract theory with its own special concepts and definitions (particularly of transfinite ordinals) is automatically applicable to actual physical events ***as they occur in time***. Bell assigns to an infinite sequence of discrete past events the order-type ω^* associated with the set of negative integers "in its natural ordering", that is the set ***terminating*** in —1. He does not consider how a sequence of events of this order-type could actually be produced. In fact, the only way in which we can define the infinite set of negative integers is by ***beginning*** with —1, but this does not correspond to the order in which the events that we may wish to associate with them occur in time. Since the set of order-type ω^* is non-constructible, there is no reason for assuming that it could represent an infinite sequence of past events. An essential feature of ***any*** past sequence of events is that it has actually occurred and been ***completed***, so that it could, in principle, have been experienced ***in toto***, or 'lived through'. Hence, if an infinite sequence of past events were possible it must be a completed infinity, which means that events must have occurred in it that were separated from the present by an infinite number of intermediate events. This is incompatible with the properties of a sequence of order-type ω^*, but contrary to what is sometimes alleged it is ***not*** assumed that there must have been a first event in the sequence. Similarly, if an infinite sequence of future events could be 'lived through', events would eventually occur that are separated from the present by an infinite number of intermediate events. Bell sees no 'logical' objection to this happening and therefore regards it as an *assumption* to assert that all future events (in sequence) are finitely accessible from an initial one, although he does not explain how in practice (*i.e.* in time) a future sequence of actual events could ever cease to be finite and become infinite. This can never happen, because in the real world actual sequences must be constructible. Consequently, although the number of future discrete events in sequence that could actually be experienced may be unbounded, it will always remain finite. (Craig uses the term 'indefinite' in this case.) This conclusion has a direct bearing on the question of whether an infinite past sequence is possible. For since, as we have seen, events must have occurred in it that were separated from the present by an infinite number of intermediate events, the present would be in the infinite future of these events and so could never occur. To avoid this contradiction, we can only conclude that any past sequence of discrete events must be finite.

In his Appendix 2, Craig criticises some of the objections that have been advanced against Kant's thesis (of his first antinomy) that there was a beginning of the universe, and he makes the comment that "The weakness of these objections is quite amazing, considering the stature of their proponents". He is careful to point out that Kant was arguing not for a beginning of time itself but for a beginning of the universe *in* time, since he conceived time in the Newtonian sense. This is even more evident in the antithesis where Kant speaks of time existing before the beginning of the world. If, as most philosophers and scientists now believe, time is inseparable from the universe, a beginning of the universe must imply a beginning of time, but as Craig points out there is no absurdity in that.

Some thirty pages of the second part of the book are devoted to an account of the current evidence from physical cosmology that points to a definite beginning of the universe. Although not an expert in this field, Craig provides an accurate summary apart from one or two minor slips. On page 111, he uses the term "Einstein-de Sitter model", where "de Sitter model" is clearly intended. (The latter depends on the cosmical constant Λ, whereas the former does not.) Also, on page 116, he says that only if the acceleration has been constant would the 'big-bang' be exactly one Hubble time ago, whereas this is the case only if the acceleration is *zero*.

The second stage of the author's defence of the *kalām* cosmological argument is discussed in two short chapters. In the first he argues that everything that begins to exist must have been caused, *i.e.* generated, by something else. He does not provide a detailed defence of this view, partly because he thinks that it is intuitively obvious, for the idea that anything could pop into existence uncaused seems to him absurd. In the following chapter he argues that Kant's antithesis (of his first antinomy) so far from disproving that there was a beginning of the universe actually provides 'a dramatic illumination of the nature of the cause of the universe'. For the antithesis asks why the universe began to exist when it did instead of existing from all eternity. Craig believes that this conundrum can best be answered not by adopting the relational concept of time, in which case the question does not arise, but by the Islamic principle of determination. According to this, when two different states of affairs are equally possible, 'the realisation of one rather than the other must be the result of the action of a personal agent who freely chooses one rather than the other'. However, whether this Creator is 'omniscient, good, perfect and so forth' is a question that the author does not consider.

Whatever view is taken of the author's treatment of the problems that he does consider, the reader should find much to interest and stimulate him in this lucidly written and carefully argued book.

G. J. WHITROW
Imperial College, London

RESCHER, N. [1979]: *Cognitive Systematisation*. Blackwells. £9.50. Pp. 211.

The questions with which Rescher is concerned are: why is it important to develop our knowledge of the world *systematically*? How does systematisation help us in the pursuit of truth? What are the most important methods of systematisation, and how are they to be validated? Is the systematicity of our knowledge simply epistemological, or also ontological? And what limits might there be to the extent to which we can systematise our knowledge? As readers of *The Coherence Theory of Truth* ([1973]) might expect, Rescher takes a quite radical coherentist line, in the spirit of Bradley or T. H. Green: systematisation is not just a way of organising our knowledge, but is actually a criterion of what counts as knowledge.

Knowledge is systematic to the extent that it conforms to certain cognitive norms: simplicity, regularity, uniformity, comprehensiveness, unity, harmony, and economy (which may, however, conflict); and the systematicity of a body of knowledge is a hallmark of its being scientific. Scientific inquiry aims at the whole truth about the world, and thus eventually to achieve a unified, that is, systematic, account of all its aspects. Systematisation serves three 'rational' (*i.e.*, cognitive, not purely psychological) functions: the systematicity of our putative knowledge makes the world intelligible to us; it organises our knowledge; *and it constitutes its verification.* The more systematic our beliefs, the less the extent of our ignorance and error; indeed, according to Rescher's 'Hegelian inversion', systematicity is itself a test of truth.

There are two main modes of systematisation, the foundationalist and the coherentist. The foundationalist (or 'Euclidean') model is of a hierarchy in which a 'self-evident' or 'incorrigible' basis supports, deductively or inductively, the rest of our knowledge. But this model is subject to dangerous tensions, for the basis must be at the same time secure and rich in content, or else the derivation must be at the same time truth-preserving and content-increasing. Gödel's theorem, furthermore, has shown the inadequacy of foundationalism as a model even of our mathematical knowledge. On the coherentist (or 'network') model, by contrast, there is no hierarchy, and hence no class of basic beliefs subject to the conflicting requirements of security and richness, but rather an egalitarian system of mutual support deriving from the inferential connections among our beliefs; and since as a result of the 'Hegelian inversion' the truths concerned are constituted by the system, the possibility of Gödelian incompleteness is excluded. The network model avoids the dichotomies of basic versus derivative knowledge, of observation statements versus laws; it does not require an input of known truths, but only of presumptively true but defeasible 'data' (which need not even be consistent); for it can select among its own data. The coherentist can accept the traditional definition of knowledge as justified, true belief; but his standards of justification or plausibility are given by the methodological rules for the construction of adequate explanations, the cognitive norms of systematisation itself. And these norms are also precisely the guiding standards of inductive reasoning, with, *e.g.*, its preference for uniformity; so that coherentism shows induction to be rationally warranted as a standard of plausibility (though not of truth).

But by what standards are coherentist standards themselves to be legitimated? The Wheel Argument shows that we cannot hope to validate our cognitive methods by appeal to the truth of their results; rather, it is the methods themselves that must somehow be warranted. This, Rescher argues, can be done by showing, at a theoretical level, that coherentist methods are self-correcting and self-substantiating, and, at a practical level, that they work well, that they enable us to intervene successfully in the world. Of these, it is practical success which is primary; and this enables Rescher to offer an evolutionary account of the growth of knowledge, according to which 'our systematized cognition fits the world for the same reasons that our eating habits do' (p. 107), both being the result of evolutionary selection, and to explain away non-scientific cognitive methods, such as astrology or numerology, as 'defunct' (p. 110). Rescher regards systematisation as a *desideratum* for our knowledge rather than as a feature of the world itself; it is possible, after all, to give a systematic description of a chaotic situation. Nevertheless, he argues, some degree of system in the world is a causally necessary condition of the possibility of rational inquiry; and granted that the world is at least sufficiently ordered to permit of rational investigation, evolution can explain why science is as successful as it is: 'It is no more a miracle that the human mind can understand the world than that the human eye can see it' (p. 123).

But still, though the world is orderly enough for science to be possible, it may be that there are, nevertheless, limits to systematisation, limits of incompleteness, disconnectedness or inconsistency. Systematicity, Rescher believes, is an *ideal*, not necessarily fully realisable. On the foundationalist model, explanation proceeds by means of subsumption under laws, and therefore can never be completed; on the coherentist model, by contrast, explanation consists of fitting a piece of information into a mutually supportive network of pieces of information, so that completeness is, in a certain sense and in principle, possible. But scientific progress always brings with it the emergence of new concepts, and hence the raising of new questions; so that it is unrealistic to hope that science could ever reach the stage where it could give 'ultimate explanations', explanations that could not be improved upon. It is also possible that our (putative) knowledge should be 'inconsequent', or disconnected, falling into distinct compartments that cannot be synthesised. And incompleteness and inconsequence together can issue in inconsistency; we may have several partial perspectives on an overall set of facts, such that the extrapolation of any of them to cover the whole range results in an inconsistency with the others. Incompleteness may, as Rescher puts it, 'exact its price not simply in *ignorance* . . . but in actual *inconsistency*' (p. 170). We can, of course, always avoid inconsistency by dropping some of our beliefs; but this may not always be the best policy, since we want to have knowledge, and not just to avoid error. We *can* always avoid inconsistency, but sometimes it is more rational to accept it. The consequences need not be disastrous; inconsistencies can be condoned if they are localised by the adoption of a logic in which a contradiction does not imply everything whatsoever.

In his last chapter Rescher turns to 'metasystematics', the systematic study of cognitive systems, and specifically to questions concerning the classification of the sciences. As science progresses, it tends towards greater specialisation; but this, Rescher argues, need not threaten the unity of science, which can, for all its specialisation, still aspire to discover uniform explanatory principles.

Rescher's presentation is austere and schematic; the ideas presented, however, are rich in applications and broad in scope. It would have been very useful to have had more comparisons, examples, and applications: for example, a discussion of the similarities and differences between Rescher's 'perspectivalism' and the apparently more relativist conceptual pluralism recently defended by Goodman,[1] or a more thorough account of how coherentism is supposed to avoid the embarrassments of a sharp dichotomy between theoretical and observation statements, or a comparison of Rescher's with Quine's or Campbell's evolutionism,[2] and so on. I should particularly have liked a more sustained presentation of the proposed *rationale* for condoning inconsistencies, because it promises both to illuminate and eventually to undermine the motivation for epistemological nihilism of the kind defended by Feyerabend, and to give relevance logic substantial epistemological application.

It would also have been desirable to have had rather more sustained and careful argument for some of the major theses which Rescher presents somewhat casually: for example, his rather perfunctory defense of the 'Hegelian inversion', in particular, is scarcely adequate in view of the weight which is subsequently placed on it; or again, his stress on practical applicability as a test of cognitive methods seems to take insufficient account of the problems involved in explaining what it means to say, or how we tell, that some application has indeed been 'successful'; and his dismissal of non-scientific methods as having failed in the evolutionary struggle leaves one wondering whether astrology, say, would become methodologically respectable if it became sufficiently fashionable.

Nevertheless, no reader—at least, no reader not incorrigibly prejudiced against idealism, coherentism, or pragmatism—could fail to find something illuminating, useful, or provocative in this book. And I am sure that many readers, like myself, will look forward to the next two volumes of Rescher's projected 'Pragmatic Epistemology' trilogy.

SUSAN HAACK
University of Warwick

REFERENCES

Goodman, N. [1972]: 'The way the world is', in *Problems and Projects*. Hackett.
Goodman, N. [1978]: *Ways of World Making*. Harvester.
Rescher, N. [1973]: *The Coherence Theory of Truth*. Oxford: U.P.
Rescher, N. [1977]: *Methodological Pragmatism*. Blackwells.

---

[1] Goodman, N. [1972] and [1978].
[2] But *cf.* Rescher [1977] for a comparison with Popper's evolutionism.

TORRETTI, ROBERTO [1978]: ***Philosophy of Geometry from Riemann to Poincaré.*** Dordrecht: D. Reidel. Pp. 358+100 p.p. notes, appendices and references. DFl 105.

Professor Torretti has written an illuminating, but not incisive, book. Within the limits he has set himself, he has covered the field in a full, fair and competent way. After a preliminary chapter on the Greek and Renaissance backgrounds, he devotes one chapter to the development of non-Euclidean geometries, the differential geometry of Gauss and Riemann, projective geometry and Klein's *Erlangen* programme, one chapter to Helmholtz's approach and the beginning of modern axiomatics up to Hilbert and Frege's criticisms of him, and a final chapter to the Empiricists, especially Mill and Mach, and their critics, to Russell, and lastly to Poincaré.

Torretti treats each thinker sympathetically, judiciously, and, with the exception, perhaps, of Russell, fairly. He uses standard modern terminology to expound their doctrines and, where necessary, to point out the flaws in their reasoning. Although the subject matter is difficult, and his authorities often obscure, his style is lucid, and his presentation a model of clarity, which other writers on geometry would do well to imitate. The reader who wants to know what Bolyai or Lie or Ueberweg or Calinon thought, and to have a sensible opinion about their thoughts without having to rethink his way through them, will always in future turn to Torretti and will find himself well served.

Torretti is rather hard on Russell. On page 317 he quotes from Russell's *Foundations of Geometry* (pp. 158):

> "What, I may be asked, is there about a thoroughly non-congruent Geometry, more impossible than this Geometry on the egg? The answer is obvious. The geometry of non-congruent surfaces is only possible by the use of infinitesimals, and in the infinitesimal all surfaces become plane. The fundamental formula, that for the length of an infinitesimal arc, is only obtained on the assumption that such an arc may be treated as a straight line, and that Euclidean Plane Geometry may be applied in the immediate neighbourhood of any point. If we had not our Euclidean measure, which could be moved without distortion, we should have no method of comparing small arcs in different places, and the Geometry of non-congruent surfaces would break down. Thus the axiom of Free Mobility, as regards three dimensional space, is necessarily implied and presupposed in the Geometry of non-congruent surfaces; the possibility of the latter, therefore, is a dependent and derivative possibility, and can form no argument against the a priori necessity of congruence as the test of equality."

and criticises it as containing "a gross misunderstanding of the fundamentals of Gauss' and Riemann's differential geometry", because the latter does not depend on embedding an n-dimensional manifold in an $(n+1)$ dimensional Euclidean space. But Russell's argument does not make any such assumption. Although his example—the surface of an egg—is, indeed, embedded, in a 3-dimensional, and at least almost Euclidean, space, his argument has nothing to do with the embedding, and rests solely on the fact that in differential geometry we assume, as indeed we do, that the space is *locally* Euclidean. Whether Russell's argument is a good one—whether our assumption of space's being at least locally Euclidean goes any way to establishing the pre-eminence of Euclidean geometry—is a question on which opinions differ: I myself think it is; others are not persuaded. But that is the argument which should be considered and, if need be, criticised—not an argument which Russell does not put forward, and which can be extracted from his example only be misreading it as being not merely an example intended to illuminate the reader but as a canonical exposition of the concept of curvature.

Torretti's dispassionate exposition falters slightly when he deals with the axiomatic approaches of Pasch, Peano and Hilbert. Here, alone in the work, I had the impression of having too much recent work being pushed at me slightly undigestedly. Section 3.22 is too long to fit into the basically historical order of exposition, too short to be intelligible to any reader who has not studied logic to a fairly advanced level. The account of categoricity (pp. 198–9) is difficult to follow, and the reader is left unclear whether it was a concept which seemed important at first but then petered out, or whether there is something which can still be salvaged and which he should try to master. In a quite different way, the reader may feel himself left up in the air in the last section of the last chapter section 4.4.6, on Poincaré's account of dimensionality.

Torretti does not maintain any philosophical thesis. Although on occasion he comments on his authorities with insight, he has no sustained theme. He is concerned with the history of the philosophy of geometry after the manner of a historian, not in the way Oxford philosophers pursue the subject. His approach enables him to treat his authors each on their own terms, but occasionally a more schematic approach would have been helpful, and sometimes a philosophical moral cries out to be drawn. Thus he draws attention to the assumption John Wallis substituted for Euclid's fifth postulate, and, on page 207 to Delboeuf's reformulation "The mutual independence of shape and size is the first postulate of geometry", but he passes by without comment (p. 178) Lie's discovery, and leaves in a footnote (n. 32, p. 415) the exegesis of Poincaré's subsequent argument, that the Euclidean group is distinguished from the two groups that leave invariant hyperbolic and elliptical geometry respectively, in that only in the Euclidean group are translations and rotations independent. It is an argument, perhaps not a conclusive one, but still quite a weighty one, for Euclidean geometry that, metaphorically speaking, it gives more degrees of freedom than its competitors: only in Euclidean geometry can shape and size vary independently, only in Euclidean geometry is a rotation quite different from a translation. Neither of these considerations is conclusive, and in the period after the close of Torretti's study Einstein urged us to trade in the simplicity of Euclidean or any other maximally symmetric geometry for the more unified explanatory power of the theory of General Relativity. That represents a radically different philosophy of geometry, in which geometry is part of physics and it is the function of geometry to explain phenomena. In Torretti's period geometry was thought of quite differently. Its function was not to explain but to provide a schema of description. Once it was realised that one could deny Euclid's fifth postulate without contradicting oneself, the problem of geometry was to discover why Euclidean geometry was to be preferred, if indeed it was to be preferred above the others, and one of its features philosophers were feeling towards was that of being more featureless, more flexible, more accommodating of more varieties of possible experience, than any other. It is a pity that Torretti does not bring this point out more explicitly.

Torretti's account of the development of geometry invites comparison with *The Development of Logic* by William and Martha Kneale. There are many points of similarity. It fails to equal the Kneale's work in philosophical penetration or synoptic vision—but then that is a high standard to be judged by, and Torretti's aim is different. He has made a valuable contribution to the history of ideas, and one which philosophers will often have occasion to use, and to be grateful for.

J. R. LUCAS
Merton College, Oxford University

EDITORIAL BOARD

R. B. Braithwaite
G. Buchdahl
L. J. Cohen
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
W. C. Kneale
S. Körner
J. L. Mackie
D. J. O'Connor
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow
J. H. Woodger

EDITORIAL ADVISORY PANEL

The Editors are helped by members of the Editorial Board, members of the Committee of the British Society for the Philosophy of Science and by an Advisory Panel which includes:

Peter Clark
Jon Dorling
Arthur Fine
Michael Gardner
Ian Hacking
Colin Howson
Noretta Koertge
David Miller
Alan Musgrave
Vernon Pratt
Ian Roxburgh
Geoffrey Sampson
Peter Urbach
Elie Zahar

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1980-81

President: Mr G. Buchdahl
President-Elect: Professor C. W. Kilmister
Hon. Secretary: Mr G. Hunt
Hon. Treasurer: Mr J. Powers
Members: Dr D. A. Gillies, Dr C. Phillips, Dr M. L. G. Redhead, Professor Y. Wilks

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all interested persons on payment of the annual subscription. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £2.50 for U.K. residents and £1.25 for overseas residents ($2.50 if resident in U.S.A. or Canada). Those who wish to join should write to the Hon. Secretary, Mr G. Hunt, Department of Philosophy, University of Warwick, Coventry CV4 7AL.

Brit. J. Phil. Sci. Vol. 31 No. 4, December 1980

The British Journal for the Philosophy of Science

Contents

15 MAY 1981

Articles

317–328 PETER URBACH Social Propensities

329–358 JOHN P. MAYBERRY A New *Begriffsschrift* (II)

Review Articles

359–375 HARVEY SIEGEL Objectivity, Rationality, Incommensurability, and More (review of T. S. Kuhn: *The Essential Tension*)

375–384 EVAN JOBE Reichenbach's Theory of Reasonable Assertion (review of H. Reichenbach: *Laws, Modalities and Counterfactuals*)

Reviews

385–416 NORETTA KOERTGE on Paul Feyerabend: *Science in a Free Society*; HANS SLUGA on Frederick Gregory: *Scientific Materialism in Nineteenth Century Germany*; IAN JARVIE on C. Hookway and P. Pettit (*eds*): *Action and Interpretation*; RICHARD SWINBURNE on Wesley Salmon (*ed*): *Hans Reichenbach: Logical Empiricist*; MICHAEL REDHEAD on D. Huff and O. Prewett (*eds*): *The Nature of the Physical Universe*; GERALD WHITROW on W. L. Craig: *The Kalam Cosmological Argument*; SUSAN HAACK on Nicholas Rescher: *Cognitive Systematisation*; J. L. LUCAS on Roberto Torretti: *Philosophy of Geometry from Riemann to Poincaré.*

Announcements

Printed in Great Britain at Aberdeen University Press